Market-oriented Two-stage Reactive Power Regulation for Large-scale Distributed Photovoltaic Entities

Qiangang Jia, Wenshu Jiao, Sijie Chen, Jian Ping, Zheng Yan, and Haitao Sun

***Abstract*—Distributed photovoltaic (PV) entities can be coordinated to provide reactive power for voltage regulation in distribution networks. However, integrating large-scale distributed PV entities into reactive power optimization makes it difficult to balance the individual benefit of each PV entity with the overall economic efficiency of the system. To address this challenge, we propose a market-oriented two-stage reactive power regulation method. At the first stage, a long-term multi-layer reactive power capacity market is created to incentivize each PV entity to provide reactive power capacity, while ensuring their financial interests are guaranteed. At the second stage, a real-time multi-layer reactive power dispatch mechanism is introduced to manage the reactive power generation of distributed PV entities, prioritizing the dispatch of lower-cost PV entities to maximize system-wide economic efficiency. Simulation results based on a real Finnish radial distribution network demonstrate the effectiveness of the proposed method in optimizing reactive power for large-scale distributed PV entities.**

***Index Terms*—Distribution network, market, distributed photovoltaic (PV), voltage regulation, reactive power regulation.**

NOMENCLATURE

A. Indexes

g	Index of bus with photovoltaic (PV) entities
G	Total number of buses with PV entities
i, j, k	Indexes of bus
m	Index of PV entity
M	Total number of PV entities
n	Index of critical bus
N	Total number of critical buses
t	Index of time step
T	Limit of time steps in an episode
Z	Limit of training episode

B. Parameters

γ_g	Intercept of aggregated bidding curve at bus g
δ	Discount factor
ε_l	Lower limit of exploration rate
ε_t	Exploration rate
θ^O	Parameter in critic network O
$\theta^{O'}$	Parameter in target critic network O'
θ^μ	Parameter in actor network μ
$\theta^{\mu'}$	Parameter in target actor network μ'
λ_g	Market clearing price at bus g
π_g	Unit annualized cost of reactive power capacity at bus g
$\pi_{g,m}$	Unit annualized cost of reactive power capacity of PV m at bus g
ρ_g	Cost changing rate of aggregated cost curve at bus g
τ	Learning speed of network parameters
Φ_j	Child bus set of bus j
χ	Decreasing rate of exploration rate
a_t	Action in deep deterministic policy gradient (DDPG) algorithm at time step t
$C_{g,m}$	Investment cost to be recovered of PV m at bus g
$E_{g,m}$	Annualized cost to be recovered of PV m at bus g
$E_{g,m,\text{active}}$	Expected annual revenue from selling active power of PV m at bus g

Manuscript received: November 5, 2024; revised: February 8, 2025; accepted: May 12, 2025. Date of CrossCheck: May 12, 2025. Date of online publication: June 6, 2025.

This work was supported by Natural Science Foundation of Henan (No. 252300421523), National Natural Science Foundation of China (No. 52422705), and Foundation of Key Laboratory of Cleaner Intelligent Control on Coal & Electricity, Ministry of Education, China (No. CICCE202409).

Q. Jia is with the School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: jiaqiangang@zzu.edu.cn).

W. Jiao is with the School of Electrical Engineering, Shandong University, Jinan 250100, China (e-mail: jknone@163.com).

S. Chen (corresponding author) and Z. Yan are with the Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China, and they are also with the Shanghai Non-Carbon Energy Conversion and Utilization Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: sijie. chen@sjtu. edu. cn; yanz@sjtu.edu.cn).

J. Ping is with the College of Smart Energy, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: ppjj1994@sjtu.edu.cn).

H. Sun is with the College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China, and he is also with the Key Laboratory of Cleaner Intelligent Control on Coal & Electricity, Ministry of Education, Taiyuan 030024, China (e-mail: sunhaitao@tyut.edu.cn).

DOI: 10.35833/MPCE.2024.001191

$E_{g,m,\text{reactive}}$	Annual cost of reactive power capacity of PV m at bus g
F	Minibatch size
I_{ij}	Current from bus i to bus j
J	Expected long-term reward under current policy
L	Mean square error in training process
$P_{g,\text{PV}}$	Active power output setting value of bus g
$P_{g,\text{PV,upper}}$	Upper historical active power generated by bus g
$P_{g,m,\text{PV,rated}}$	Rated active power of PV m at bus g
P_{ij}	Active power from bus i to bus j
P_{jk}	Active power from bus j to bus k
$P_{j,\text{load,ini}}$	Benchmark active power consumption value of bus j
$P_{j,\text{load}}$	Active power consumption at bus j
$Q_{g,\text{PV}}$	Reactive power output obligation of bus g
$Q_{g,m,\text{PV}}$	Reactive power output obligation of PV m at bus g
$Q_{g,\text{PV,cap}}$	Reactive power capacity awarded at bus g
$Q^{*}_{g,\text{PV,cap}}$	Available capacity that does not affect active power output at bus g
$Q_{g,m,\text{PV,cap}}$	Reactive power capacity awarded to PV m at bus g
$Q^{*}_{g,m,\text{PV,cap}}$	Available capacity that does not affect active power output of PV m at bus g
$Q_{g,m,\text{PV,rated}}$	Rated reactive power of PV m at bus g
$Q_{g,\text{PV,require}}$	Reactive power requirement at bus g
Q_{ij}	Reactive power from bus i to bus j
Q_{jk}	Reactive power from bus j to bus k
$Q_{j,\text{load}}$	Reactive power consumption at bus j
$Q_{j,\text{load,ini}}$	Benchmark reactive power consumption value of bus j
r_{ij}	Resistance from bus i to bus j
$r_{\text{reward},t}$	Reward in DDPG algorithm at time step t
$R_{n,\text{punish}}$	Normalized punishment at bus n
s_t	State in DDPG algorithm at time step t
$S_{g,m,\text{PV,rated}}$	Rated apparent power of PV m at bus g
U	Data replay buffer in DDPG training
V_i	Voltage magnitude at bus i
V_j	Voltage magnitude at bus j
$V_{j,\text{ini}}$	Benchmark voltage magnitude at bus j
V_n	Voltage magnitude of critical bus n
$V_{\min}$	Lower limit of voltage magnitude
$V_{\max}$	Upper limit of voltage magnitude
W_g	Profit of awarded bus g
x_{ij}	Reactance from bus i to bus j
y_t	Value estimated from target critic network at time step t
$Y_{g,m}$	Remaining service life of PV m at bus g

I. Introduction

HIGH penetration of distributed renewable energy leads to significant over-voltage issues caused by reverse power flow in distribution networks. To enhance voltage safety, the distribution system operator (DSO) has the ability to regulate the reactive power output of inverter-based devices, especially distributed photovoltaic (PV) entities [1].

Existing research primarily focuses on the effectiveness of voltage control methods by forcibly regulating the reactive power output of distributed PV entities. Analytical methods [2]-[4] derive reactive power optimization plans by obtaining global information on the parameters of PV resources and the states of distribution networks. In recent years, data-driven methods have been widely applied to reactive power optimization. These methods typically leverage machine learning algorithms to enable real-time adjustment of reactive power output from PV and other distributed energy resources (DERs) under incomplete information [5], [6]. Reference [7] proposed a reactive power optimization framework for Energy Internet during voltage sags based on multiagent deep reinforcement learning. Reference [8] proposed a real-time two-time-scale voltage regulation method with a soft open point to solve the reactive power optimization problem. Reference [9] developed a multi-agent deep reinforcement learning algorithm to address autonomous voltage control issues. However, these methods primarily focused on mandatorily regulation without offering incentives for individual PV entity. As a result, PV entities have little motivation to actively participate in such programs.

To incentivize PV entities to provide reactive power, reactive power pricing on the power distribution side is gradually carried out [10]. Reference [11] explored the possibility of wind turbines providing reactive power and proposed a pricing method for reactive power capacity when their behavior is uncertain. Reference [12] explored the possibility of PV and energy storage systems providing reactive power and clarified a pricing method for distributed PV systems in grid-connected microgrids to provide reactive power. Reference [13] proposed a reactive power pricing mechanism based on the Vickery-Clarke-Groves auction for DERs. Reference [14] proposed a reactive power pricing model based on the location marginal price of distribution network. These methods introduced pricing principles that incentivized DERs to participate in reactive power optimization. However, existing research often focuses on nodal reactive power pricing, which struggles to account for the diverse characteristics and variations among numerous PV entities.

To sum up, the existing methods struggle to bridge the gap between the financial interests of distributed PV entities and the global voltage regulation targets of the DSO. To fill this gap, this paper proposes a two-stage multi-layer reactive power regulation method. The main contributions are as follows.

1) A long-term multi-layer reactive power capacity market is designed to incentivize PV entities to provide the reactive

power capacity. By aggregating bids and decomposing market results, the financial interests of distributed PV entities are guaranteed, even with multiple participants involved.

2) A real-time multi-layer reactive power dispatch mechanism is designed to allocate reactive power for voltage regulation at the lowest cost. Using the deep deterministic policy gradient (DDPG) algorithm, reactive power generation responsibilities are quickly assigned to each bus, and then redistributed to the low-cost PV entities through linear rules to enhance efficiency.

The remainder of this paper is organized as follows. Section II is the preview of proposed two-stage multi-layer regulation method. Section III details the long-term multi-layer reactive power capacity market. Section IV explains the real-time multi-layer reactive power dispatch mechanism. A case study is performed in Section V. Section VI concludes the paper.

II. Preview of Proposed Two-stage Multi-layer Regulation Method

A. Framework of Multi-layer Interaction

The framework of the multi-layer interaction is depicted in Fig. 1.

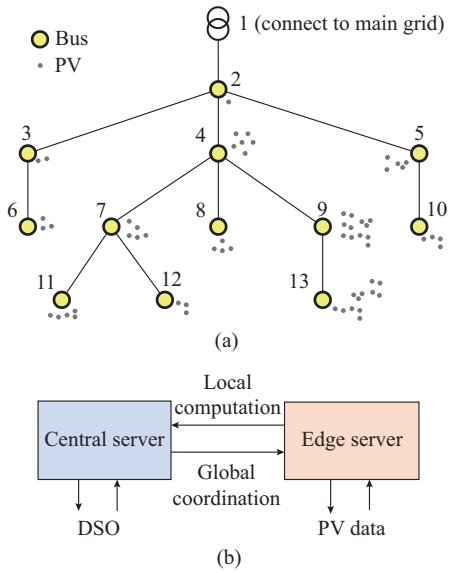

Fig. 1. Framework of multi-layer interaction. (a) Typical radial power distribution system with PV entities. (b) Data interaction between central and edge servers in reactive power regulation.

Figure 1(a) shows a typical radial power distribution system containing massive PV entities. Figure 1(b) shows the data interaction between central and edge servers in reactive power regulation. The number of edge servers corresponds to the number of buses, and they are responsible for processing local PV data. A single central server is responsible for communication and data exchange with the edge servers. This central server is typically neutral, which helps protect PV data from being directly exposed to the DSO.

B. Overview of Two-stage Mechanism

There are two stages: long-term reactive power capacity configuration and real-time reactive power dispatch. The relationship of the two stages is shown in Fig. 2.

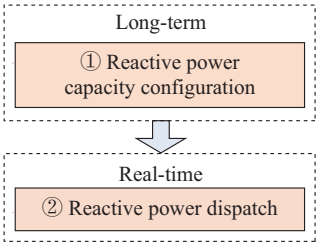

Fig. 2. Relationship of proposed two stages.

1) Stage 1: Long-term Reactive Power Capacity Configuration

In this stage, different PV entities are assumed to submit their bids for reactive power capacity to the central server through edge servers. The central server then executes the market clearing algorithm to determine the reactive power capacity price and allocate the awarded capacity to each bus as well as the corresponding PV entities. PV entities will earn revenues in this configuration while the DSO obtains sufficient regulation capability. Compared with conventional reactive power pricing methods, our method accounts for the differences among the numerous distributed PV entities, ensuring the economic benefits of individual participants.

2) Stage 2: Real-time Reactive Power Dispatch

In this stage, the DSO aims to prevent system voltage violations by regulating the reactive power output of the PV entities involved at stage 1. The central and edge servers implement a DDPG algorithm and linear rules to achieve real-time reactive power dispatch. The awarded capacity of lower-cost PV entities from stage 1 will serve as the reactive power dispatch limits at stage 2. Compared with traditional reactive power dispatch methods, our method explicitly defines the dispatch costs of different PV entities. By prioritizing dispatch based on cost from lowest to highest, this method enhances the economic efficiency of global reactive power optimization.

Note that we focus on the optimization of light assets like PV entities at the above two stages under specific conditions, i.e., the distribution network topology is fixed, and the discrete operational statuses of heavy assets like switched capacitor banks, step voltage regulators, and transformers with on-load tap changers (OLTCs) are predefined.

III. Long-term Multi-layer Reactive Power Capacity Market

The interaction of different servers and participants in the long-term multi-layer reactive power capacity market is shown in Fig. 3, which consists of six steps.

Step 1: PV entity submits the bid of reactive power capacity to the local edge servers in the nearest bus.

Step 2: after all data from PV entities are collected, the edge servers aggregate the bids and submit them to the central server.

Step 3: DSO submits the network parameters to the central server.

Step 4: central server solves the market clearing problem according to the aggregated bids and the network parameters.

Step 5: central server releases the market clearing results to the DSO and edge servers.

Step 6: edge servers decompose the market clearing results to the PV entities.

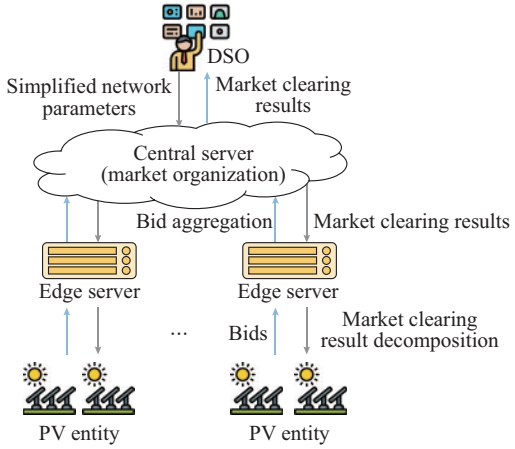

Fig. 3. Interaction of different servers and participants in long-term multi-layer reactive power capacity market.

The details of the critical points, e. g., bid aggregation, market clearing result decomposition, and market organization are analyzed in the following parts.

A. Bid Aggregation and Market Clearing Result Decomposition

The annualized cost to be recovered is as follows:

$$E_{g,m} = C_{g,m}/Y_{g,m} \tag{1}$$

Given the expected annual revenue from selling active power $E_{g,m,\text{active}}$:

1) If $E_{g,m,\text{active}} \geq E_{g,m}$, it indicates that distributed PV entity m can recover the investment costs solely through selling active power, so the annual cost of reactive power capacity is zero.

2) If $E_{g,m,\text{active}} < E_{g,m}$, it indicates that distributed PV entity m cannot recover the investment costs solely through selling active power, so the annual cost of reactive power capacity can be expressed as:

$$E_{g,m,\text{reactive}} = E_{g,m} - E_{g,m,\text{active}} \tag{2}$$

As shown in Fig. 4, the rated active, reactive, and apparent power of PV m at bus g satisfies the following relationship.

$$Q^2_{g,m,\text{PV,rated}} = S^2_{g,m,\text{PV,rated}} - P^2_{g,m,\text{PV,rated}} \tag{3}$$

Assume that the available reactive power capacity that does not affect the active power output is as follows:

$$0 \leq Q^*_{g,m,\text{PV,cap}} \leq \left| Q_{g,m,\text{PV,rated}} \right| \tag{4}$$

Calculate the unit annualized cost $\pi_{g,m}$ of reactive power capacity as:

$$\pi_{g,m} = E_{g,m,\text{reactive}}/Q^*_{g,m,\text{PV,cap}} \tag{5}$$

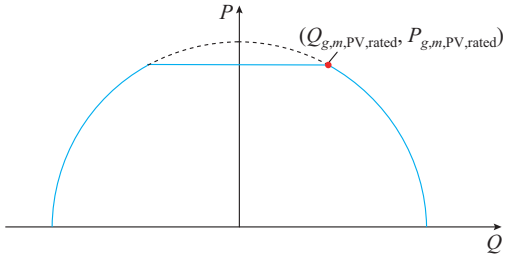

Fig. 4. Relationship of active, reactive, and apparent power.

1) Original Bids

As shown in Fig. 5, PV entities calculates the available reactive power capacity amount-price pairs.

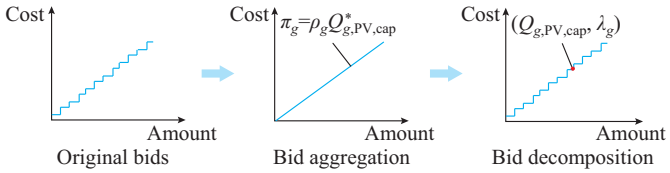

Fig. 5. Illustration of original bids, bid aggregation, and bid decomposition processes.

The original bids of PV entities at bus g are listed as follows:

$$\left\{ Q^*_{g,1,\text{PV,cap}}, Q^*_{g,2,\text{PV,cap}}, ..., Q^*_{g,m,\text{PV,cap}}, ..., Q^*_{g,M,\text{PV,cap}} \right\} \tag{6}$$

$$\left\{ \pi_{g,1}, \pi_{g,2}, ..., \pi_{g,m}, ..., \pi_{g,M} \right\} \tag{7}$$

2) Bid Aggregation

To simplify the market clearing problem, a line based on the regression methods is taken to replace the amount-price pairs in the stepwise bidding curve.

$$\pi_g = \rho_g Q^*_{g,\text{PV,cap}} + \gamma_g \tag{8}$$

Note that the line will pass through the origin if PV entities are diverse enough.

$$\gamma_g = 0 \tag{9}$$

Then, the slope of the curve and the available capacity of all PV entities at bus g will be submitted for market clearing.

3) Market Clearing Result Decomposition

As shown in Fig. 5, within a particular bus g, PV entities whose bids are lower than the market clearing price are awarded the reactive power capacity until the total awarded amount of bus g is reached. The awarded reactive power capacities of all PV entities at bus g are cleared at price λ_g. Since the reactive power value varies at each bus, the pay-as-bid theory [15] better reflects the differences among nodes compared with the uniform clearing mechanism based on marginal pricing.

$$\lambda_g = \rho_g Q_{g,\text{PV,cap}} \tag{10}$$

B. Market Organization

This subsection aims to determine the awarded capacities of different buses. The market is cleared to minimize the total cost of purchasing reactive power capacity from all buses with PV entities.

$$\min_{Q_{g,\mathrm{PV,cap}}} \sum_{g=1}^{G} \rho_g Q_{g,\mathrm{PV,cap}}^2 \tag{11}$$

To reduce the complexity of the market clearing problem, the simplified power flow model [16] is used in the optimization constraints.

$$P_{ij} = \sum_{k \in \Phi_j} P_{jk} - P_{g,\mathrm{PV,upper}}\Big|_{g=j} \tag{12}$$

$$Q_{ij} = \sum_{k \in \Phi_j} Q_{jk} - Q_{g,\mathrm{PV,require}}\Big|_{g=j} \tag{13}$$

Since high penetration of DERs often leads to over-voltage issues due to reverse power flow, the extreme situation is considered, i.e., each bus with PV entities generates an upper value of active power, and the power consumption of each bus is zero. Note that the upper historical active power value is a boundary condition, it could be calculated based on information such as historical solar irradiance.

The following constraints are set since the value of reactive power capacity is positive.

$$Q_{g,\mathrm{PV,require}} \le Q_{g,\mathrm{PV,cap}} \tag{14}$$

$$-Q_{g,\mathrm{PV,require}} \le Q_{g,\mathrm{PV,cap}} \tag{15}$$

Similarly, we use the simplified voltage drop equations in [16] to describe the relationship between the power flow and voltage magnitudes at different buses.

$$V_j = V_i - P_{ij} r_{ij} - Q_{ij} x_{ij} \tag{16}$$

The awarded reactive power capacity $Q_{g,\mathrm{PV,cap}}$ of bus g with PV entities should not exceed $Q^*_{g,\mathrm{PV,cap}}$.

$$Q_{g,\mathrm{PV,cap}} \le Q^*_{g,\mathrm{PV,cap}} \tag{17}$$

The voltage magnitudes of critical buses should be restricted to a predefined safe range.

$$V_{\min} \le V_n \le V_{\max} \tag{18}$$

The profit of awarded bus g with PV entities can be calculated after the market clearing process.

$$W_g = 0.5 \rho_g Q_{g,\mathrm{PV,cap}}^2 \tag{19}$$

IV. Real-time Multi-layer Reactive Power Dispatch Mechanism

The interaction of the edge servers and the central server in the real-time multi-layer reactive power dispatch mechanism is shown in Fig. 6.

The reactive power dispatch limits are bounded by awarded reactive power capacities at stage 1. The reactive power dispatch process consists of four steps, which are given as follows.

Step 1: DSO trains the DDPG network and sends the parameters of the trained actor network to the central server.

Step 2: edge servers measure the voltage magnitude and sends it to the central server.

Step 3: the central server sets the reactive power output plan of each bus according to voltage magnitudes by actor network, then sends it to the edge servers.

Step 4: edge servers set the reactive power dispatch plan to each PV entity according to their proportions of awarded reactive power capacities.

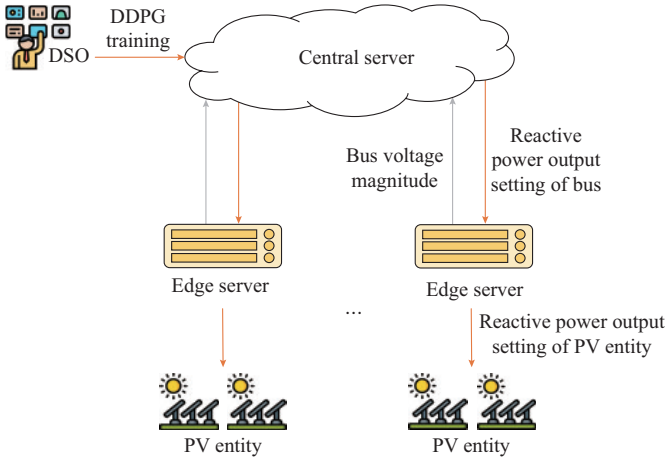

Fig. 6. Interaction of different servers and participants in real-time multi-layer reactive power dispatch mechanism.

The DDPG training, reactive power output setting of bus, and reactive power output setting of PV entity are analyzed in detail as follows.

A. DDPG Training

The reactive power dispatch mechanism aims to maintain voltage deviations at critical buses within a predefined safe range. The power flow and voltage drop equations are given as follows, which cannot be simplified any further because of accuracy.

$$P_{ij} = \sum_{k \in \Phi_j} P_{jk} + P_{j,\mathrm{load}} - P_{g,\mathrm{PV}}\Big|_{g=j} + r_{ij} I_{ij} \tag{20}$$

$$Q_{ij} = \sum_{k \in \Phi_j} Q_{jk} + Q_{j,\mathrm{load}} - Q_{g,\mathrm{PV}}\Big|_{g=j} + x_{ij} I_{ij} \tag{21}$$

$$P_{j,\mathrm{load}} = P_{j,\mathrm{load,ini}} \left(\frac{V_j}{V_{j,\mathrm{ini}}} \right)^2 \tag{22}$$

$$Q_{j,\mathrm{load}} = Q_{j,\mathrm{load.ini}} \left(\frac{V_j}{V_{j,\mathrm{ini}}} \right)^2 \tag{23}$$

$$V_j^2 = V_i^2 - 2P_{ij} r_{ij} - 2Q_{ij} x_{ij} + \left(r_{ij}^2 + x_{ij}^2 \right) I_{ij} \tag{24}$$

$$I_{ij} V_i^2 = P_{ij}^2 + Q_{ij}^2 \tag{25}$$

The DDPG algorithm is trained in a simulated environment, as illustrated in Fig. 7.

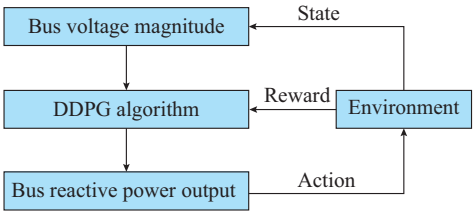

Fig. 7. Illustration of DDPG algorithm.

The objective of the DDPG algorithm is to generate optimal dispatch plans at the bus level. The key characteristics of the DDPG algorithm are state, action, and reward.

State s: the voltage magnitude values of N critical buses, consisting of N-dimension variables.

$$s=\left\{V_1, V_2, \ldots, V_n, \ldots, V_N\right\} \tag{26}$$

Action a: the reactive power output of G buses with PV entities, consisting of G-dimension variables.

$$a=\left\{a_1, a_2, \ldots, a_g, \ldots, a_G\right\} \quad a_g \in \left[-Q_{g,\mathrm{PV,cap}}, Q_{g,\mathrm{PV,cap}}\right] \tag{27}$$

Reward r_{reward}: if the voltage magnitude of a critical bus n is out of the predefined safe zone, a normalized punishment in $[-1, 0]$ is given. Thus, the total reward of N critical buses is as follows.

$$r_{\mathrm{reward}}=\begin{cases}0 & V_n \in [0.95, 1.05]\ \mathrm{p.u.} \\ \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} R_{n,\mathrm{punish}} & V_n \notin [0.95, 1.05]\ \mathrm{p.u.}\end{cases} \tag{28}$$

The learning principles of the DDPG algorithm are based on the theory in [17]. It contains a critic network, a target critic network, an actor network, and a target actor network.

1) The critic network is set to estimate the value function.

2) The actor network determines the actions based on the current states.

The rule for updating the critic network is to minimize the mean square error L.

$$L\left(\theta^O\right)=\frac{1}{F}\sum_t\left(y_t-O\left(s_t, a_t|\theta^O\right)\right)^2 \tag{29}$$

$$y_t=r_{\mathrm{reward},t}+\delta O'\left(s_{t+1}, \mu'\left(s_{t+1}|\theta^{\mu'}\right)|\theta^{O'}\right) \tag{30}$$

The rule for updating the actor network is based on the chain rule.

$$\nabla_{\theta^\mu} J \approx \frac{1}{F}\sum_t \nabla_a O\left(s, a|\theta^O\right)\Big|_{s=s_t, a=\mu\left(s_t\right)} \nabla_{\theta^\mu}\mu\left(s|\theta^\mu\right)\Big|_{s_t} \tag{31}$$

The updated rules of the target critic network and the target actor network are as follows.

$$\theta^{O'} \leftarrow \tau\theta^O+(1-\tau)\theta^{O'} \tag{32}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu+(1-\tau)\theta^{\mu'} \tag{33}$$

The training process of the DDPG algorithm is shown in Algorithm 1. In each episode, the DDPG algorithm generates an action based on the current state and receives a reward. This process is called a Markov decision process. Note that the exploration rate in the training process is decreasing.

$$\varepsilon_{t+1}=\begin{cases}\chi\varepsilon_t & \varepsilon_t \geq \varepsilon_l \\ \varepsilon_l & \varepsilon_t < \varepsilon_l\end{cases} \tag{34}$$

B. Reactive Power Output Setting of Bus

The well-trained DDPG algorithm is sent to the central server for implementation in the target environment. The implementation of the DDPG algorithm is shown in Algorithm 2.

C. Reactive Power Output Setting of PV Entity

Edge servers allocate the reactive power output to PV entities based on linear rules, i.e., the proportion of awarded reactive power capacity of different PV entities at each bus. The awarded PV entity will be responsible for the reactive power output obligation as follows:

Algorithm 1

1. **Input**: critical bus voltage level
2. **Output**: reactive power output of the buses with PV entities
3. Initialize critic and actor networks
4. Initialize target networks
5. Initialize data replay buffer U
6. **for** $episode=1{:}Z$ do
7. Initialize the power flow solver and calculate the voltage magnitude of critical buses
8. Receive the initial observation state s_0
9. **for** time step $t=1{:}T$ do
10. Select actions a_t according to the current policy and state
11. Execute the action a_t, receive reward r_t, and turn to new state s_{t+1}
12. Store transition $\left(s_t, a_t, r_{\mathrm{reward},t}, s_{t+1}\right)$ in data replay buffer U
13. Sample a random minibatch of F transitions from U
14. Update critic and actor networks
15. Update target networks
16. **if** the voltage magnitudes at all critical buses are in a tolerance range, break
17. **end if**
18. **end for**
19. **end for**

Algorithm 2

1. **Input**: voltage magnitudes of all critical buses
2. **Output**: reactive power output of the buses with PV entities
3. Utilize the trained networks
4. Receive the observed state
5. Select actions according to the state

$$Q_{g,m,\mathrm{PV}}=\frac{Q_{g,m,\mathrm{PV,cap}}}{Q_{g,\mathrm{PV,cap}}}Q_{g,\mathrm{PV}} \tag{35}$$

By scheduling the reactive power output responsibility to awarded low-cost PV resources, the overall reactive power optimization cost is minimized.

V. Case Study

The case studies are simulated by a server with an Intel Core i9 CPU. The topology of a real Finnish radial distribution system is shown in Fig. 8 [18]. Buses 5, 9, 12, 17, 22, 27, 30, 33, and 38 are designated as critical buses. Buses 5, 7, 20, 24, 31, 36, 38, and 39 host PV entities (assuming each bus hosts 24 PV entities, totaling 192 PV entities). The predefined safe zone of voltage magnitude is set to be [0.95, 1.05]p.u. (the base value of active/reactive power is 10 MW/Mvar, and the base value of voltage magnitude is 20 kV). The resistance and reactance of each transmission line are both set to be 0.01 p.u..

A. Clearing Results of Reactive Power Capacity Market

The simulation is performed based on MATLAB R2022b. We use the Gurobi solver and YALMIP compiler. The time to solve this analytical market clearing problem is about 1.6 s, mainly depending on the performance of the hardware and the optimization solvers.

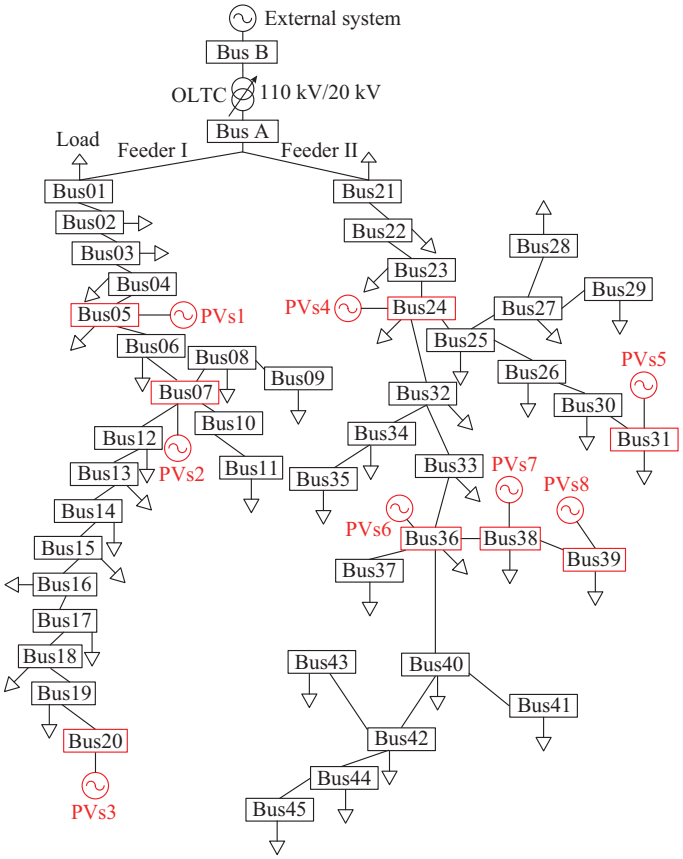

Fig. 8. Topology of a real Finnish radial distribution system.

A typical bus consisting of 24 PV entities is initially configured, with unit annualized costs and available capacities generated randomly. Then, it is expanded to other buses by multiplying slopes and quantities by random coefficients. The bids from a typical bus with PV entities are illustrated in Fig. 9, where each discrete point represents the midpoint value of the corresponding amount-price pair. The slope is 2.693 \$/Mvar^2, and the total available capacity of this bus is 8.794 Mvar. The random multipliers for buses with PV entities are shown in Table I.

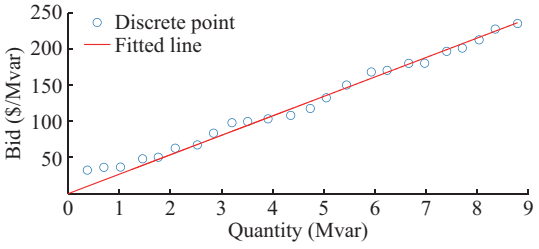

Fig. 9. Bids of a typical bus containing PV entities.

TABLE I
RANDOM MULTIPLIERS FOR BUSES WITH PV ENTITIES

Bus No.	Multiplier	Bus No.	Multiplier
5	1.36	31	2.20
7	2.66	36	1.80
20	2.61	38	2.79
24	2.79	39	3.45

Table II provides the market clearing results of each bus with PV entities (the voltage magnitude of reference bus A is regulated to 1.00 p.u. by OLTC). Assume that the upper active power generated by each bus with PV entities is set to be the same value of 10 MW.

TABLE II
MARKET CLEARING RESULTS OF BUSES WITH PV ENTITIES

Bus No.	Slope (\$/Mvar^2)	Price (\$/Mvar)	$Q^*_{g,\mathrm{PV,cap}}$ (Mvar)	$Q_{g,\mathrm{PV,cap}}$ (Mvar)	Profit (\$)
5	3.657	30.30	11.944	8.284	125.50
7	7.153	42.42	23.358	5.930	125.77
20	7.027	64.59	22.947	9.192	296.85
24	7.502	39.66	24.499	5.287	104.83
31	5.916	43.73	19.319	7.392	161.64
36	4.847	44.56	15.830	9.192	204.78
38	7.514	69.07	24.538	9.192	317.43
39	9.303	84.15	30.379	9.045	380.57

Bus 39 has the highest slope, while bus 5 has the lowest slope. Buses 20, 36, and 38 are awarded the largest reactive power capacity, while bus 24 is awarded the smallest reactive power capacity. The slopes and positions of a bus determine the market clearing results. A lower slope and a more critical position lead to a larger awarded reactive power capacity. Buses with PV entities gain considerable profits: bus 39 achieves the highest profit of \$380.57, whereas bus 24 achieves the lowest profit of \$104.83. The total profit of all buses with PV entities is \$1717.37 through the reactive power capacity market.

In market clearing result decomposition, all PV entities at the same bus will be settled at the market clearing price of that bus (30.30 to 84.15 \$/Mvar). All PV entities in the bus with bids lower than this price will be awarded until the total reactive power capacity (5.29 to 9.19 Mvar) is met.

For example, only 6 PV entities are awarded reactive power capacities at bus 24, as shown in Table III. Besides, the PV entities with lower costs will achieve higher profits in this process. The market mechanism guarantees the economic interests of each PV entity and improves the utilization of PV equipment. The situation is similar for PV entities at other buses, so further details are omitted.

TABLE III
DATA OF AWARDED PV ENTITIES AT BUS 24

Awarded PV entity	Reactive power capacity (Mvar)	Revenue (\$)
1	1.06	42.13
2	0.90	35.76
3	0.91	35.94
4	1.21	47.87
5	0.86	33.92
6	0.35	14.07

To demonstrate the influence of OLTC at reference bus A, the voltage magnitude of reference bus A V_{A} is changed by OLTC to different values. The simulation results in Fig. 10

show that a lower voltage magnitude of reference bus A leads to a lower purchasing amount of reactive power capacity.

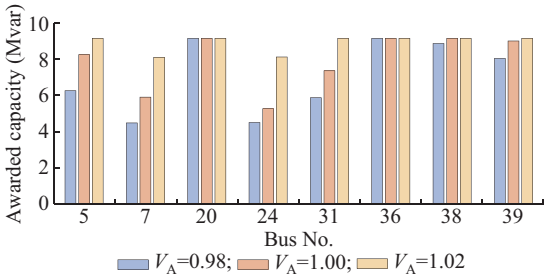

Fig. 10. Market clearing results under different voltage magnitudes of reference bus A.

B. Real-time Reactive Power Dispatch Results

The simulation is performed based on Python 3.8.3. We use the Anaconda environment with the PyTorch structure. We assume that each bus with PV entities randomly generates active power within 10 MW, while each bus with loads randomly generates active and reactive power consumptions within 0.2 MW and 0.1 Mvar at the benchmark voltage of 1.0 p.u.. The voltage magnitude of reference bus A is set to be 1.0 p. u.. Note that the neural networks should be retrained if the voltage magnitude of reference bus A is changed by OLTC.

In the DDPG networks, the number of neurons in a hidden layer is 32. The initial exploration rate is 0.6, and the lower limit ε_l is 0.05. The memory limit of the replay buffer is 500, and the minibatch size is 32. The decreasing rate is 0.99995. The learning rate of actor and critic networks is 0.001. The DDPG algorithm is trained using 2000 randomly generated episodes, each with a limitation of 20 iterations. The number of total training iterations is 40000. The convergence criterion in the training process is that the total reward is larger than 0.8. As illustrated in Fig. 11, the average total normalized reward increases during the training process.

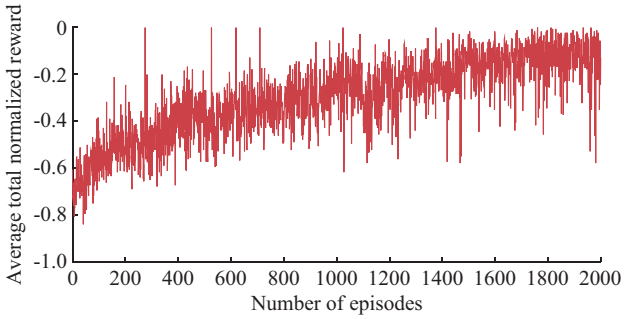

Fig. 11. Training process of DDPG algorithm.

The statistical characteristics of voltage magnitudes at critical buses during 50 tests are depicted in Fig. 12. It can be observed that the proposed method reduces the voltage magnitude at critical buses and restricts the voltage fluctuation into the predefined safe range.

Figure 13 shows the statistical characteristics of reactive power dispatch results of each bus with PV entities using the violin plot.

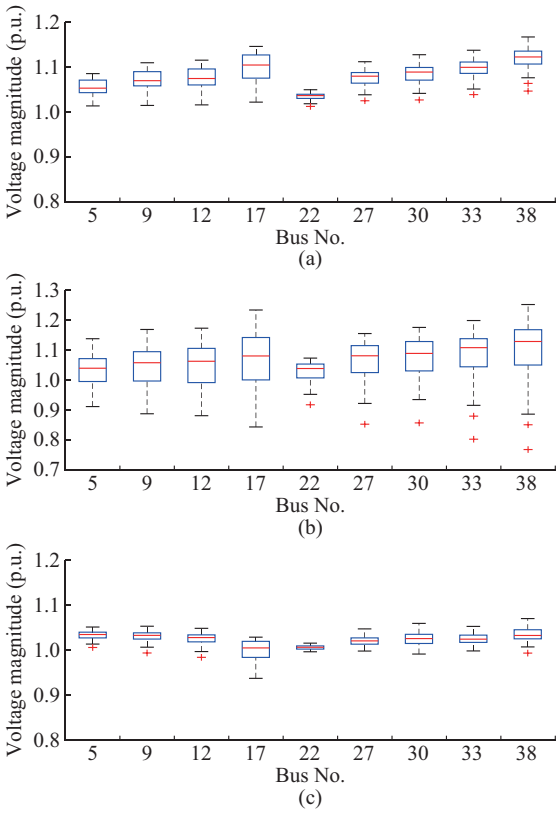

Fig. 12. Statistical characteristics of voltage magnitudes at critical buses during tests. (a) Results without regulation (scenario 1). (b) Results when dispatch amounts are randomly selected within awarded reactive power capacities of each bus as baseline (scenario 2). (c) Results under proposed method (scenario 3).

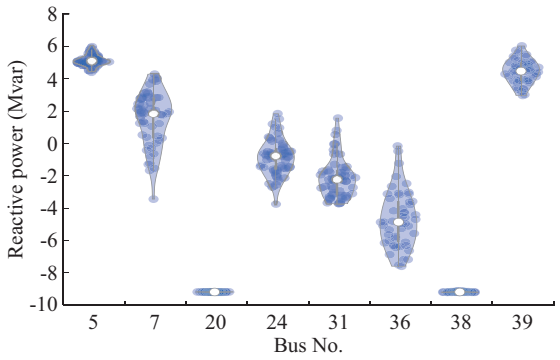

Fig. 13. Statistical characteristics of reactive power dispatch results of each bus with PV entities.

Figure 13 combines the features of a box plot and a density plot to visualize the distribution shape of reactive power output of each bus. Table IV shows the average reactive power dispatch results of each bus with PV entities. Buses 5, 7, and 39 are dispatched to generate positive reactive power, while buses 20, 24, 31, 36, and 38 are dispatched to generate negative reactive power. They are coordinated to maintain voltage safety in the distribution system.

To show the reactive power dispatch results of PV entities at a bus during the testing process, bus 24 is analyzed in detail, as shown in Fig. 14. The 6 PV entities awarded in the reactive power capacity market share the dispatch obligation

in proportion to their awarded capacity.

TABLE IV
AVERAGE REACTIVE POWER DISPATCH RESULTS OF BUSES WITH PV ENTITIES

Bus No.	Reactive power (Mvar)	Bus No.	Reactive power (Mvar)
5	5.092	31	–2.102
7	1.461	36	–4.693
20	–9.176	38	–9.168
24	–0.850	39	4.434

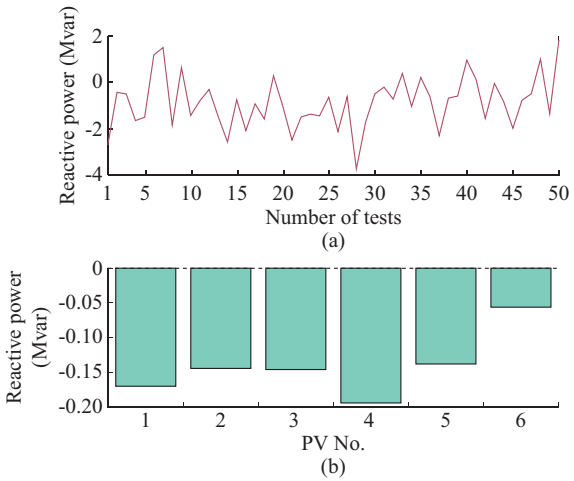

Fig. 14. Reactive power dispatch results of PV entities at bus 24. (a) Reactive power dispatch result of bus 24. (b) Average value of reactive power dispatch result sharing at bus 24.

The voltage magnitude samples out of the predefined safe range are summarized in Table V. The voltage magnitude of reference bus A is changed by OLTC to different values. Through the proposed method, the samples out of the predefined safe zone are all within 5% in different scenarios. When the voltage magnitude of reference bus A is 1.00 p.u., the samples out of the predefined safe zone are only 15 in scenario 3, accounting for 3.33% of the total 450 (50×9) samples.

TABLE V
VOLTAGE MAGNITUDE SAMPLES OUT OF PREDEFINED SAFE RANGE

Scenario	Number of samples			Percentage (%)		
	0.98 p.u.	1.00 p.u.	1.02 p.u.	0.98 p.u.	1.00 p.u.	1.02 p.u.
1	257	354	409	57.11	78.67	90.89
2	291	299	317	64.67	66.44	70.44
3	18	15	20	4.00	3.33	4.44

The time consumption of optimization and communication [19], [20] for the reactive power dispatch process is analyzed as follows. The communication latency between the central server and the local server is nearly zero in the 5G communication network. The reactive power dispatch for each bus by the DDPG network is calculated within 5.1 ms, while the time required for dispatching reactive power to each PV entity via linear affine is negligible. Therefore, the total time consumption for a dispatch signal from the central server to the PV entities is about 5 ms. Note that the inference time of the DDPG algorithm heavily depends on the computing platform. If a more powerful CPU or general processing unit (GPU) is utilized, the inference time could be further reduced to 1 ms or less, making the system even more suitable for fast-response applications.

VI. CONCLUSION

This paper proposes a market-oriented two-stage reactive power regulation method to coordinate the DSO and distributed PV entities. Firstly, the long-term multi-layer reactive power capacity market configures reactive power capacity while ensuring the benefits of each PV entity. Case studies show that through the long-term multi-layer reactive power capacity market, all buses with PV entities achieve a profit of \$1717.37. Secondly, the real-time multi-layer reactive power dispatch mechanism ensures the rapid allocation of reactive power obligations from the DSO to the distributed PV entities in the most cost-effective manner. Case studies demonstrate that in a 450-sample voltage magnitude simulation, the proposed method achieves a reliability rate exceeding 95% by optimally dispatching lower-cost PV entities. The benefits of each PV entity and the DSO are guaranteed.

In the future, distribution network reconfiguration can be considered to coordinate PV entities to realize more effective voltage regulation. Meanwhile, some advanced reinforcement learning algorithms, e.g., soft actor-critic and twin delayed DDPG, can be adopted in reactive power dispatch. Besides, distributed dispatch schemes can improve privacy protection in the regulation process.

REFERENCES

[1] W. Jiao, J. Chen, Q. Wu *et al.*, "Distributed coordinated voltage control for distribution networks with DG and OLTC based on MPC and gradient projection," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 680-690, Jan. 2022.
[2] W. Ma, W. Wang, Z. Chen *et al.*, "Voltage regulation methods for active distribution networks considering the reactive power optimization of substations," *Applied Energy*, vol. 284, p. 116347, Feb. 2021.
[3] C. Zhang, Y. Xu, Y. Wang *et al.*, "Three-stage hierarchically-coordinated voltage/var control based on PV inverters considering distribution network voltage stability," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 868-881, Apr. 2022.
[4] Y. Ju, Z. Zhang, W. Wu *et al.*, "A bi-level consensus ADMM-based fully distributed inverter-based volt/var control method for active distribution networks," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 476-487, Jan. 2022.
[5] S. Hou, E. M. Salazar, P. Palensky *et al.*, "A mix-integer programming based deep reinforcement learning framework for optimal dispatch of energy storage system in distribution networks," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 2, pp. 597-608, Mar. 2025.
[6] D. S. Pacheco-Cherrez, D. Guillen, J. C. Mayo-Maldonado *et al.*, "Data-driven optimal voltage performance index tracking in active distribution networks," *IEEE Transactions on Smart Grid*, vol. 15, no. 5, pp. 4804-4818, Sept. 2024.
[7] S. Xu and S. Guo, "Distributed reactive power optimization for energy Internet via multiagent deep reinforcement learning with graph attention networks," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 6, pp. 8696-8706, Jun. 2024.
[8] M. Xiong, X. Yang, Y. Zhang *et al.*, "Reactive power optimization in active distribution systems with soft open points based on deep reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109601, Jan. 2024.
[9] S. Wang, J. Duan, D. Shi *et al.*, "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning,"

IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 4644-4654, Nov. 2020.
[10] A. Potter, R. Haider, G. Ferro *et al*., "A reactive power market for the future grid," *Advances in Applied Energy*, vol. 9, p. 100114, Feb. 2023.
[11] A. C. Rueda-Medina and A. Padilha-Feltrin, "Distributed generators as providers of reactive power support: a market approach," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 490-502, Feb. 2013.
[12] O. Gandhi, C. D. Rodríguez-Gallegos, W. Zhang *et al*., "Economic and technical analysis of reactive power provision from distributed energy resources in microgrids," *Applied Energy*, vol. 210, pp. 827-841, Jan. 2018.
[13] M. Jiang, Q. Guo, H. Sun *et al*., "Leverage reactive power ancillary service under high penetration of renewable energies: an incentive-compatible obligation-based market mechanism," *IEEE Transactions on Power Systems*, vol. 37, no. 4, pp. 2919-2933, Jul. 2022.
[14] J. Stekli, L. Bai, and U. Cali, "Pricing for reactive power and ancillary services in distribution electricity markets," in *Proceedings of 2021 IEEE PES Innovative Smart Grid Technologies Conference*, Washington DC, USA, Feb. 2021, pp. 1-5.
[15] G. Federico and D. M. Rahman, "Bidding in an electricity pay-as-bid auction," *SSRN Electronic Journal*, vol. 24, no. 2, pp. 175-211, Sept. 2003.
[16] Y. Song, Y. Zheng, T. Liu *et al*., "A new formulation of distribution network reconfiguration for reducing the voltage volatility induced by distributed generation," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 496-507, Jan. 2020.
[17] T. P. Lillicrap, J. J. Hunt, A. Pritzel *et al*. (2015, Sept.). Continuous control with deep reinforcement learning. [Online]. Available: http://arxiv.org/abs/1509.02971
[18] A. Kulmala, S. Repo, and P. Jarventausta, "Coordinated voltage control in distribution networks including several distributed energy resources," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 2010-2020, Jul. 2014.
[19] L.-N. Liu, G.-H. Yang, and S. Wasly, "Distributed predefined-time dual-mode energy management for a microgrid over event-triggered communication," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3295-3305, Mar. 2024.
[20] L.-N. Liu, G.-H. Yang, and S. Wasly, "Distributed event-triggered economic environmental resource management for islanded microgrids under DoS attacks," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 4, pp. 7158-7169, Oct. 2024.

Qiangang Jia received the B.S. degree in electrical engineering from Shandong University, Jinan, China, in 2018, and the Ph.D. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China. He is currently a Lecturer in School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou, China. His research interests include electricity market, optimization of distributed energy resources, machine learning, energy blockchain, and demand response.

Wenshu Jiao received the B.S. and M.S. degrees in electrical engineering from Shandong University, Jinan, China, in 2018 and 2021, respectively, where he is currently working toward the Ph.D. degree in electrical engineering. His research interests include renewable energy generation, active distribution network, voltage control, and distributed optimization algorithm with applications to power system.

Sijie Chen received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2009 and 2014, respectively. He is currently a tenured Associate Professor of electrical engineering with Shanghai Jiao Tong University, Shanghai, China. His research interests include energy blockchain, electricity market, and distributed optimization.

Jian Ping received the B.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2015 and 2020, respectively. He is currently an Assistant Professor with the College of Smart Energy, Shanghai Jiao Tong University. His research interests include energy blockchain, transactive energy system, and electricity market.

Zheng Yan received the B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1987 and 1991, respectively. He is currently a Professor of electrical engineering with Shanghai Jiao Tong University. His research interests include application of optimization theory to power systems, electricity market, and dynamic security assessment.

Haitao Sun received the M.Sc. degree in electrical engineering from Taiyuan University of Technology, Taiyuan, China, and the Ph.D. degree in electromechanical engineering from Ghent University, Ghent, Belgium, in 2016 and 2020, respectively. He became a Lecturer with Taiyuan University of Technology, in 2020. His research interests include electrical drive, power electronic inverter, and electrical machine.