

ORIGINAL ARTICLE

International Journal of Applied Mathematics in Control Engineering

Journal homepage: <http://www.ijamce.com>

Multi-Agent Reinforcement Learning Bidding Strategy for Virtual Power Plants Based on Regularization Mechanism

Zeyu Li^{1*} | Junhua Gu^{2*}

¹College of Electrical Engineering, Hebei University of Technology, Tianjin, Tianjin, 300401, China

²College of Electrical Engineering, Hebei University of Technology, Tianjin, Tianjin, 300401, China

Correspondence

College of Electrical Engineering, Hebei University of Technology, Tianjin, Tianjin, 300401, China
Email: 18332316621@163.com

Article Info

Article history:

Received 22 March 2026

Accepted 28 May 2026

Available online 1 June 2026

Abstract

With the rapid development of virtual power plants (VPPs) and electricity spot markets, effective bidding strategies are essential for improving market profitability and reflecting realistic price formation mechanisms. However, traditional optimization-based bidding models often rely on complete information assumptions, while existing multi-agent reinforcement learning methods may produce clearing prices that deviate from real market trends. To address these issues, this paper proposes a regularized multi-agent twin delayed deep deterministic policy gradient algorithm, denoted as RM-MATD3, for VPP bidding in electricity markets. A market trading model is first established, including bidding models for thermal units and VPPs as well as a locational marginal pricing-based market clearing model. In the proposed framework, each market participant is modeled as an agent, bidding prices are regarded as actions, clearing results are treated as observations, and market profits are used as rewards. Furthermore, a real-price-guided regularization mechanism is embedded into the actor network to constrain learned bidding strategies and improve consistency with actual electricity prices. Case studies based on the IEEE 30-bus system and real day-ahead market price data show that RM-MATD3 improves training stability, preserves strategic bidding space, and effectively reduces the deviation between simulated and actual clearing prices. The results demonstrate the feasibility and practical value of the proposed method for VPP spot market bidding.

KEYWORDSreinforcement learning, *strategic bidding*, virtual power plant

1 | INTRODUCTION

China is accelerating power system reform, and the traditional “generation following load” paradigm can no longer meet the requirements of the new-type power system. As an important form of aggregating distributed energy resources, virtual power plants (VPPs) play a key role in power balance management, optimal dispatch, and flexible response. Meanwhile, the unified national electricity market and electricity spot markets are being steadily promoted, with full coverage and continuous settlement operation expected by the end of 2025. As the scale of individual VPPs expands, their influence on clearing prices becomes

increasingly significant. Therefore, studying the game equilibrium between VPPs and other market participants is essential for understanding market price dynamics and improving bidding profitability.

Existing studies have extensively investigated the bidding strategies of generation companies and VPPs. Some works model VPPs as price takers and solve bidding models using evolutionary algorithms [1], while others construct repeated game models or bi-level bidding models to derive optimal strategies under Nash or Stackelberg equilibrium [2]. However, these studies generally assume that VPP bids do not affect market clearing results, making it difficult to capture the potential price-making power of VPPs in competitive markets.

To characterize strategic interactions, many studies further treat market participants as price makers and reformulate bi-level bidding problems into single-level models using Karush–Kuhn–Tucker (KKT) conditions and linearization techniques. This framework has been extended to VPP bidding under renewable uncertainty, spot market risks, carbon trading, segmented bidding, and coalition game scenarios [3, 4, 5]. Nevertheless, most of these studies assume that competitors' bidding strategies are known in advance and rely on complete information, which deviates from real electricity market conditions. In addition, KKT-based reformulation and linearization may introduce approximation errors and weaken nonlinear characteristics.

To overcome these limitations, multi-agent reinforcement learning (MARL), as a model-free method suitable for incomplete-information environments, has been increasingly applied to electricity market bidding [6]. Existing studies show that algorithms such as MADQN, MADDQN, MADDPG, and MATD3 can characterize strategic bidding interactions and support convergence toward market equilibria under certain conditions [7]. These methods have also been extended to continuous action spaces, multi-market coupling, coalition bidding, and electricity–carbon–green certificate trading scenarios [8, 9]. Hierarchical reinforcement learning has been introduced to reduce bidding decision complexity, but it usually increases network structure complexity and training time. Moreover, clearing prices generated by MARL-based simulations may deviate from real electricity market prices, limiting their practical profitability. Therefore, accurately characterizing market equilibrium behavior while maintaining consistency with real market price levels remains a key challenge for VPP spot market bidding.

To investigate the optimal bidding strategies of virtual power plants participating in electricity markets and the game equilibrium consistent with actual market clearing outcomes, this paper presents the market trading process, the bidding models of electricity market participants, and the clearing model of the market operator.

1.1 | Market Trading Process

The trading process of the electricity spot market forms a critical link between the bidding behavior of market participants and the price formation mechanism, and it has a direct impact on market clearing outcomes and the economic benefits of all participants. Therefore, before analyzing the bidding strategies and game equilibrium of virtual power plants, a brief description of the trading process of the day-ahead electricity spot market is provided.

(1) Prior to the bidding day, the market operator releases boundary condition information for the operating day to all market participants, including system load forecasts, renewable energy generation forecasts (such as wind and photovoltaic power), reserve requirements, and relevant system constraints.

(2) Based on the forecast information and their own operational characteristics, generation companies and electricity consumers submit quantity–price bidding curves or tables to the market operator.

(3) After collecting all submitted bids, the market operator runs the market clearing model while considering system operation constraints (e.g., transmission capacity limits and reserve requirements) to determine nodal clearing prices and generation schedules for all market participants.

(4) Through the electricity trading platform, the market operator publishes the generation and consumption schedules and the clearing prices for the operating day to all market participants, providing the basis for actual operation on the following day.

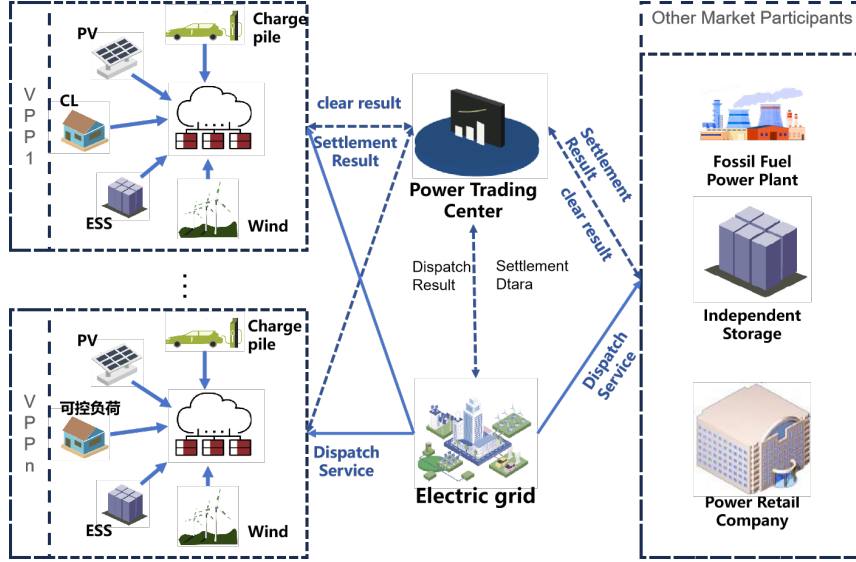


FIGURE 1 Spot Trading Framework

2 | ELECTRICITY MARKET TRADING MODEL

To investigate the optimal bidding strategies of virtual power plants participating in electricity markets and the game equilibrium consistent with actual market clearing outcomes, this paper presents the market trading process, the bidding models of electricity market participants, and the clearing model of the market operator.

2.1 | Bidding Model

Traditional thermal power units submit bids based on their generation output levels, with the objective of maximizing their profits.

$$\max W = \sum_{t=1}^T \lambda_t p_t - (ap_t^2 + bp_t + c) \quad (1)$$

where W denotes the total profit of the traditional thermal power unit; λ_t is the market clearing price at time period t ; p_t represents the awarded generation quantity of the thermal unit at time period t ; and $ap_t^2 + bp_t + c$ is the quadratic cost function of the thermal power unit. T denotes the total number of bidding periods within one day.

$$\left\{ \begin{array}{l} \max W_{vpp} = W_{wind} + W_{pv} + W_{ess} + W_{cl} \\ W_{wind} = \sum_{t=1}^T (\lambda_t - \lambda_w) p_t^{wind} \\ W_{pv} = \sum_{t=1}^T (\lambda_t - \lambda_{pv}) p_t^{pv} \\ W_{ess} = \sum_{t=1}^T (\lambda_t^{dis} p_t^{ess,dis} + \lambda_t^{ch} p_t^{ess,ch}) \\ W_{cl} = -\lambda_t p_t^{cl} \end{array} \right. \quad (2)$$

where W_{wind} denotes the revenue of wind power; p_t^{wind} is the awarded wind power output; and λ_w represents the unit generation cost of wind power. W_{pv} denotes the revenue of photovoltaic generation; p_t^{pv} is the awarded photovoltaic output; and λ_{pv} is the unit generation cost of photovoltaic power. W_{ess} denotes the revenue of the energy storage system; $p_t^{ess,dis}$ is the awarded discharging power; λ_t^{dis} is the discharging price; $p_t^{ess,ch}$ is the awarded charging power; and λ_t^{ch} is the charging price. W_{cl} denotes the revenue of controllable loads, and p_t^{cl} represents the awarded controllable load power.

2.2 | Market Clearing Model

After market participants submit their bids, the power trading center performs market clearing using security-constrained unit commitment (SCUC) and security-constrained economic dispatch (SCED) procedures. A centralized bidding mechanism is adopted, and prices are determined based on locational marginal pricing (LMP). The clearing objective is to maximize social welfare, equivalently to minimizing the total electricity purchase cost of society, and the objective function is formulated as:

$$\min \sum_{t=1}^T \left(\sum_{i=1}^n \lambda_t^{g,i} p_t^{g,i} + \sum_{j=1}^J \lambda_t^{\text{wind},j} p_t^{\text{wind},j} + \sum_{k=1}^K \lambda_t^{\text{pv},k} p_t^{\text{pv},k} + \sum_{m=1}^M \lambda_t^{\text{ess},m} p_t^{\text{ess},m} - \sum_{u=1}^U \lambda_t^{\text{cl},u} p_t^{\text{cl},u} \right) \quad (3)$$

where $p_t^{g,i}$ and $\lambda_t^{g,i}$ denote the offered power and offer price of thermal generating unit i , respectively; $p_t^{\text{wind},j}$ and $\lambda_t^{\text{wind},j}$ denote the offered power and offer price of wind generating unit j ; $p_t^{\text{pv},k}$ and $\lambda_t^{\text{pv},k}$ denote the offered power and offer price of photovoltaic unit k ; $p_t^{\text{ess},m}$ and $\lambda_t^{\text{ess},m}$ denote the offered power and offer price of energy storage unit m ; and $p_t^{\text{cl},u}$ and $\lambda_t^{\text{cl},u}$ denote the offered power and offer price of controllable load u .

Constraints:

Global power balance constraint:

$$\sum_{i=1}^I p_t^{g,i} + \sum_{j=1}^J p_t^{\text{wind},j} + \sum_{k=1}^K p_t^{\text{pv},k} + \sum_{m=1}^M p_t^{\text{ess},m} - \sum_{u=1}^U p_t^{\text{cl},u} = \sum_{b=1}^B p_t^{\text{load},b} \quad (4)$$

where B denotes the total number of buses, and $p_t^{\text{load},b}$ represents the load demand at bus b at time period t .

Thermal unit awarded power constraint:

$$0 \leq p_t^{\text{bid},g,i} \leq p_t^{g,i} \quad (5)$$

Thermal unit ramping constraint:

$$\Delta p_{\min}^{g,i} \leq p_t^{\text{bid},g,i} - p_{t-1}^{\text{bid},g,i} \leq \Delta p_{\max}^{g,i} \quad (6)$$

where $\Delta p_{\min}^{g,i}$ and $\Delta p_{\max}^{g,i}$ denote the lower and upper limits of the ramping power of the thermal unit, respectively.

Wind power output constraint:

$$0 \leq p_t^{\text{bid},\text{wind},j} \leq p_t^{\text{wind},j} \quad (7)$$

Photovoltaic power output constraint:

$$0 \leq p_t^{\text{bid},\text{pv},k} \leq p_t^{\text{pv},k} \quad (8)$$

Energy storage power constraint:

$$0 \leq p_t^{\text{bid},\text{ess},m} \leq p_t^{\text{ess},m} \quad (9)$$

Energy storage state-of-charge (SOC) constraint:

$$\text{SOC}_{\min} \leq \text{SOC}_t = \frac{p_t^{\text{ess},\text{ch},m} \eta_{\text{ch}} - p_t^{\text{ess},\text{dis},m} / \eta_{\text{ch}}}{E_{\max}} \leq \text{SOC}_{\max} \quad (10)$$

To ensure that the energy storage system remains dispatchable on the following day, the total charging and discharging energy within one day must be zero; that is, the SOC at the end of the day is equal to that at the beginning.

$$\sum_{t=1}^T \left(p_t^{\text{ess},\text{ch},m} \eta_{\text{ch}} - p_t^{\text{ess},\text{dis},m} \eta_{\text{dis}} \right) = 0 \quad (11)$$

Nodal power balance constraint:

$$P_{i_b,t} - P_{D,i_b,t} - \sum_{(i_b,j_b) \in \mathcal{K}} P_{i_b,j_b,t} = 0 \quad (12)$$

Transmission line power flow constraint:

$$\begin{cases} P_{i_b j_b, t} = \frac{\theta_{i_b, t} - \theta_{j_b, t}}{X_{i_b j_b}}, & \forall (i_b, j_b) \in \mathcal{K}, \\ -P_{i_b j_b}^{\max} \leq P_{i_b j_b, t} \leq P_{i_b j_b}^{\max}, & \forall (i_b, j_b) \in \mathcal{K}. \end{cases} \quad (13)$$

where $P_{i_b j_b, t}$ denotes the net injected power on line (i_b, j_b) at time period t ; $\theta_{i_b, t}$ denotes the voltage phase angle at bus i_b ; and $P_{i_b j_b}^{\max}$ is the maximum transmission capacity of the transmission line between buses i_b and j_b .

3 | REGULARIZED MULTI-AGENT REINFORCEMENT LEARNING ALGORITHM

To solve the non-cooperative bidding game among multiple market participants under incomplete information, this paper adopts a regularized multi-agent twin delayed deep deterministic policy gradient algorithm, denoted as RM-MATD3. In the proposed framework, each market participant is regarded as an agent, the electricity market clearing model serves as the environment, the bidding prices are defined as actions, the clearing results are taken as observations, and the market profits are used as rewards.

3.1 | MATD3-Based Bidding Framework

For an electricity market with N participants, the state space, action space, and reward set are defined as

$$S = \{o_1, o_2, \dots, o_N\} \quad (14)$$

$$o_i = \left\{ p_{w, t}^{\text{bid}}, p_{v, t}^{\text{bid}}, p_{\text{ess}, t}^{\text{bid}}, p_{\text{cl}, t}^{\text{bid}}, \lambda_t^n \right\} \quad (15)$$

$$A = \{a_1, a_2, \dots, a_N\} \quad (16)$$

$$R = \{r_1, r_2, \dots, r_N\} \quad (17)$$

where o_i , a_i , and r_i denote the observation, bidding action, and profit reward of agent i , respectively. For VPP agents, the observation includes the awarded power of wind power, photovoltaic generation, energy storage, controllable load, and the corresponding nodal electricity price. For thermal generation agents, the observation consists of the awarded generation output and nodal electricity price.

MATD3 adopts the centralized training and decentralized execution framework. The actor network generates bidding actions based on local observations, while the critic network evaluates the joint action value using global information. To reduce overestimation and improve training stability, two critic networks are introduced. The actor objective of agent i is expressed as

$$J'(\theta) = \mathbb{E}_{s, a \sim \mathcal{B}} \left[Q(s, a_1, a_2, \dots, a_N) \Big|_{a_i = \mu_i(o_i)} \right] + \delta \mathbb{E} \left[(\lambda_t^r - \mu(a_i | o_i))^2 \right] \quad (18)$$

where $\mathcal{B}$ denotes the replay buffer, $\mu_i(\cdot)$ is the actor policy of agent i , and $Q_i(\cdot)$ is the critic network.

The critic loss is given by

$$L_{\text{critic}} = \mathbb{E}_{s, a, s', r} \left[(Q_i(s, \mathbf{a}) - y)^2 \right], \quad \mathbf{a} = (a_1, \dots, a_N) \quad (19)$$

where the target value is calculated as

$$y = r + \gamma \min \left(Q_i^1(s', a'_1, \dots, a'_N), Q_i^2(s', a'_1, \dots, a'_N) \right) \Big|_{a'_i = \mu'_i(o_i)} \quad (20)$$

with γ representing the discount factor and $Q'_{i, k}$ denoting the target critic networks.

Under incomplete information, each agent cannot directly obtain the real strategies of other participants. Therefore, opponent

$$L(\phi_i^j) = -\mathbb{E}_{o_j, a_j} \left[\ln \hat{\mu}_i^j(a_j | o_j) + \lambda H(\hat{\mu}_i^j) \right] \quad (21)$$

$$L(\phi_i^j) = -\mathbb{E}_{o_j, a_j} \left[\ln \hat{\mu}_i^j(a_j | o_j) + \lambda H(\hat{\mu}_i^j) \right] \quad (21)$$

3.2 | Real-Price-Guided Regularization Mechanism

$$J'(\theta) = \mathbb{E}_{s,a \sim \mathcal{B}} \left[Q(s, a_1, a_2, \dots, a_N) \Big|_{a_i = \mu_i(o_i)} \right] + \delta \mathbb{E} \left[(\lambda_t^r - \mu(a_i \mid o_i))^2 \right] \quad (22)$$

Since the main actor objective changes during training, a dynamic regularization coefficient is adopted:

$$\delta = \min(100 \ln(1 + \beta J(\theta_i)), \delta_{\max}), \quad (23)$$

3.3 | Training Procedure

FIGURE 2 Overall Flow of the Algorithm

4 | CASE STUDY

4.1 | Parameter Settings

This paper adopts the IEEE 30-bus standard test system to conduct case studies on the proposed model and its solution algorithm. Day-ahead market clearing data publicly released by the Shanxi Electric Power Trading Center are used. Three virtual power plants (VPPs) and two thermal power units are considered. Specifically, the first VPP aggregates wind power at Bus 1, controllable loads at Bus 2, and energy storage devices at Bus 3; the second VPP aggregates wind power at Bus 8, photovoltaic generation at Bus 11, and energy storage devices at Bus 13; and the third VPP aggregates photovoltaic generation at Bus 1 and controllable loads at Bus 20. Thermal power units 1 and 2 are located at Bus 5 and Bus 11, respectively. Detailed parameter settings are provided in the Appendix. According to the current electricity market rules in Shanxi Province, the upper limit of the bidding price for market participants is set to 1.5 CNY/kWh. Algorithm parameter settings: The scheduling horizon consists of 24 time periods. The maximum number of iterations for each time period is 700, with the first 100 iterations used for random exploration and network parameter updates starting from the 101st iteration. The replay buffer size is set to 250, the discount factor γ is 0.99, and the learning rate of the neural networks is 0.01. Simulation environment: Intel Core i7-8750 CPU @ 2.20 GHz, 8 GB RAM. The software environment includes Python 3.7.16 and PyTorch 1.7.1.

4.2 | Algorithm Performance Analysis

(1) Comparison of MATD3 with MADDPG, MADDQN, and MADQN

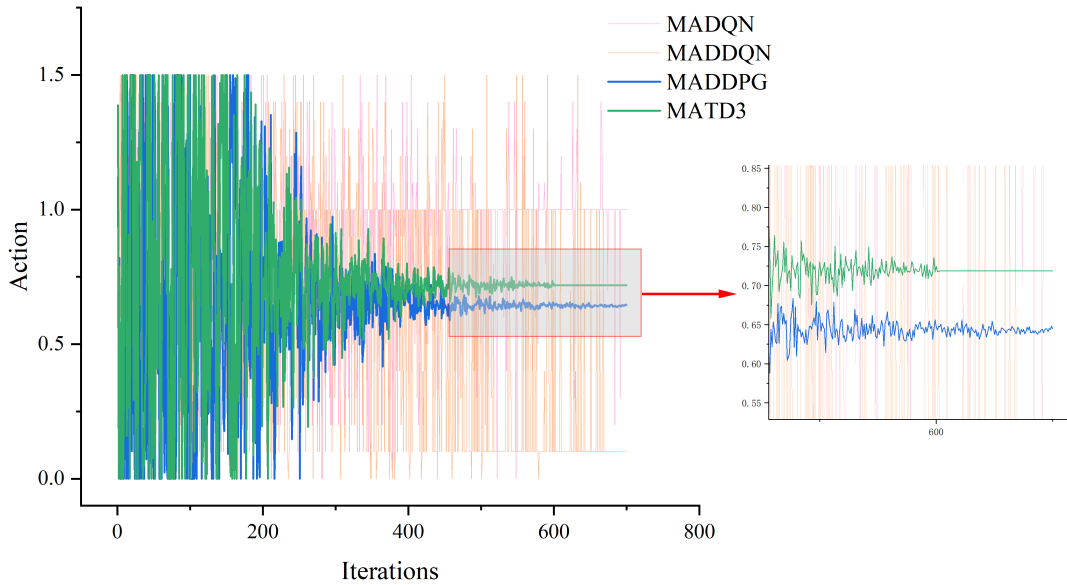


FIGURE 3 Algorithm Comparison

Fig. 3 compares the convergence speed of four multi-agent reinforcement learning methods. As shown in the figure, the two discrete-action algorithms, MADDQN and MADQN, exhibit a convergence trend within 700 iterations but fail to converge. The MADDPG algorithm approaches convergence around 700 iterations; however, it still shows noticeable fluctuations after 600 iterations. In contrast, the MATD3 algorithm demonstrates a significant reduction in fluctuations after 600 iterations, indicating more stable convergence. These results suggest that MATD3 performs better than the other methods in solving the proposed problem.

(2) Performance Comparison Between the Regularization Mechanism and the Baseline Training Strategy

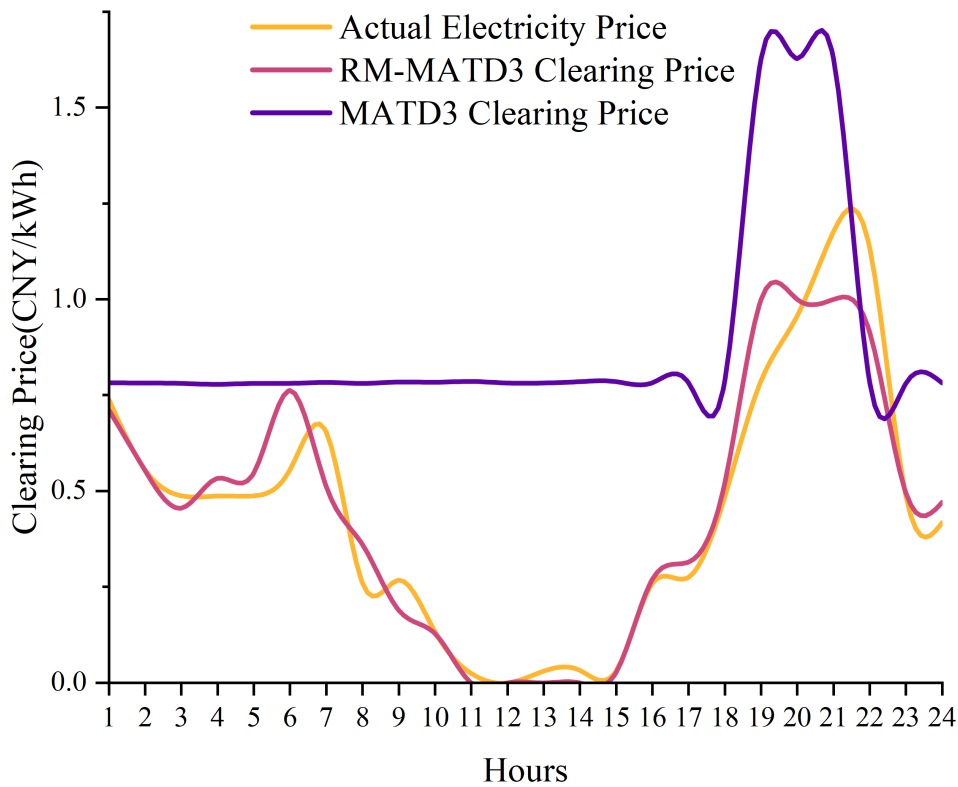


FIGURE 4 Comparison of actual electricity price, MATD3 and RM-MATD3 clearing prices

Fig. 4 compares the clearing prices obtained from the real electricity market, the conventional MATD3 algorithm, and the proposed MATD3 with an embedded regularization mechanism (regularization mechanism MATD3, RM-MATD3). As shown in the figure, the clearing prices derived from the conventional multi-agent reinforcement learning approach do not vary with changes in supply–demand conditions. For most time periods, the clearing price remains around 0.75 CNY/kWh, which deviates significantly from real market prices, resulting in an average absolute price deviation of 0.47056. Such results provide limited reference value for virtual power plant bidding decisions. After incorporating the regularization term, the clearing prices obtained by the multi-agent reinforcement learning model move closer to real market prices, although they do not exactly coincide with them. This is because a certain degree of strategic bidding space is intentionally preserved to ensure that the resulting game equilibrium remains meaningful and practically relevant.

(3) Adjustment of the Regularization Parameter

Fig. 5 illustrates the impact of the regularization coefficient on the extent to which agents' actions align with real electricity prices. As shown in the figure, when the coefficient is set to 50, the bidding actions of all agents strictly follow the real market prices, leaving almost no room for strategic competition. When the coefficient is reduced to 25, the influence of real prices on agents' actions is weakened, and a certain degree of bidding game space is released. As the coefficient further decreases, the available game space gradually increases. Table 2 presents the effect of different coefficient values on the price deviation metric.

Table 2 presents the impact on electricity price deviation, evaluated using three metrics: Mean Absolute Deviation (MAD), Symmetric Relative Deviation (SRD), and Mean Squared Deviation (MSD). The corresponding calculation formulas are given as

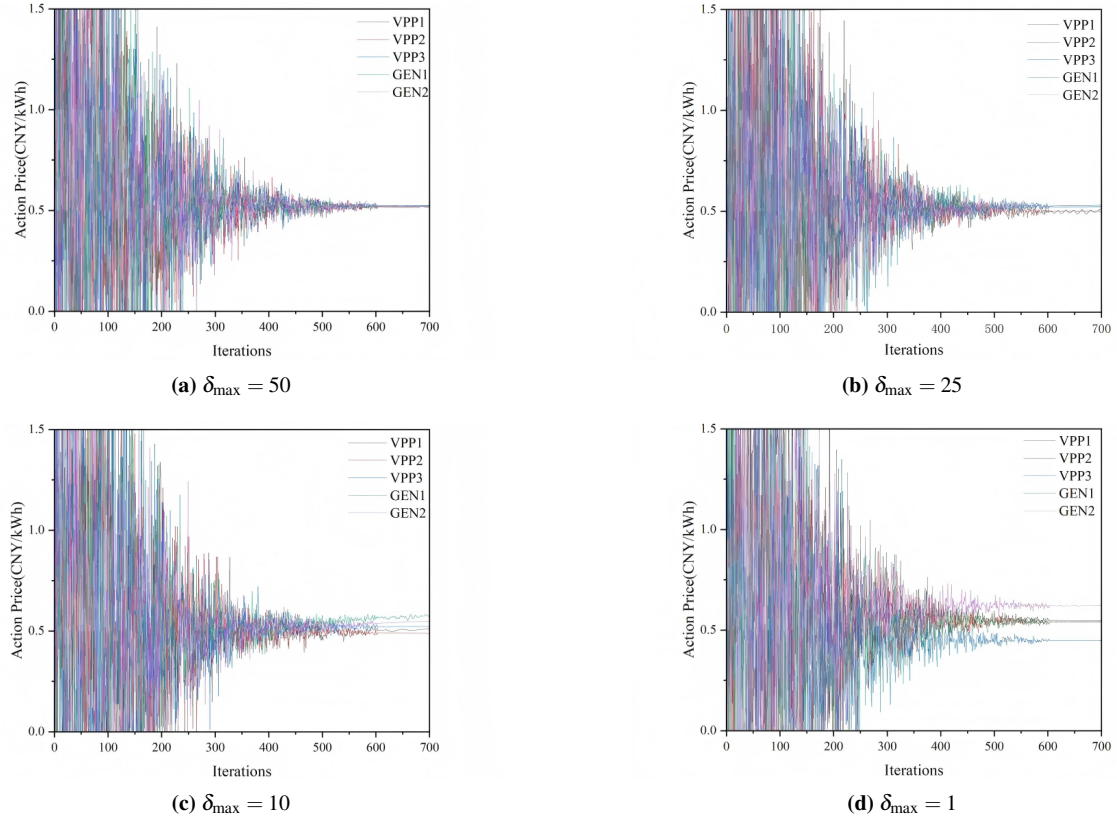


FIGURE 5 Impact of Value on Actions of Each Agent

follows:

$$\begin{aligned}
 \text{MAD} &= \frac{1}{T} \sum_{t=1}^T |\lambda_t - \lambda_t^r|, \\
 \text{SRD} &= \frac{1}{T} \sum_{t=1}^T \frac{|\lambda_t - \lambda_t^r|}{(\lambda_t^r + \lambda_t)/2}, \\
 \text{MSD} &= \frac{1}{T} \sum_{t=1}^T (\lambda_t - \lambda_t^r)^2.
 \end{aligned} \tag{24}$$

where λ_t denotes the clearing price obtained from the multi-agent reinforcement learning simulation environment, and λ_t^r represents the real market electricity price. As shown in Table 2, with an increase in the regularization coefficient, the influence of Regularization Term on the simulated clearing prices becomes stronger, and all deviation metrics decrease significantly, indicating closer alignment with real market prices. However, a certain degree of strategic bidding space should be preserved, and excessive restriction on agent actions should be avoided. Therefore, coefficient values in the range of 10–25 provide a good balance between price realism and game flexibility.

TABLE 1 Impact of δ value on electricity price deviation

δ	MAD	SRD	MSD
1	0.36	73.78%	0.31
10	0.17	62.40%	0.10
25	0.12	46.42%	0.03
50	0.07	35.25%	0.01

(4) Analysis of Bidding Outcomes and Profitability

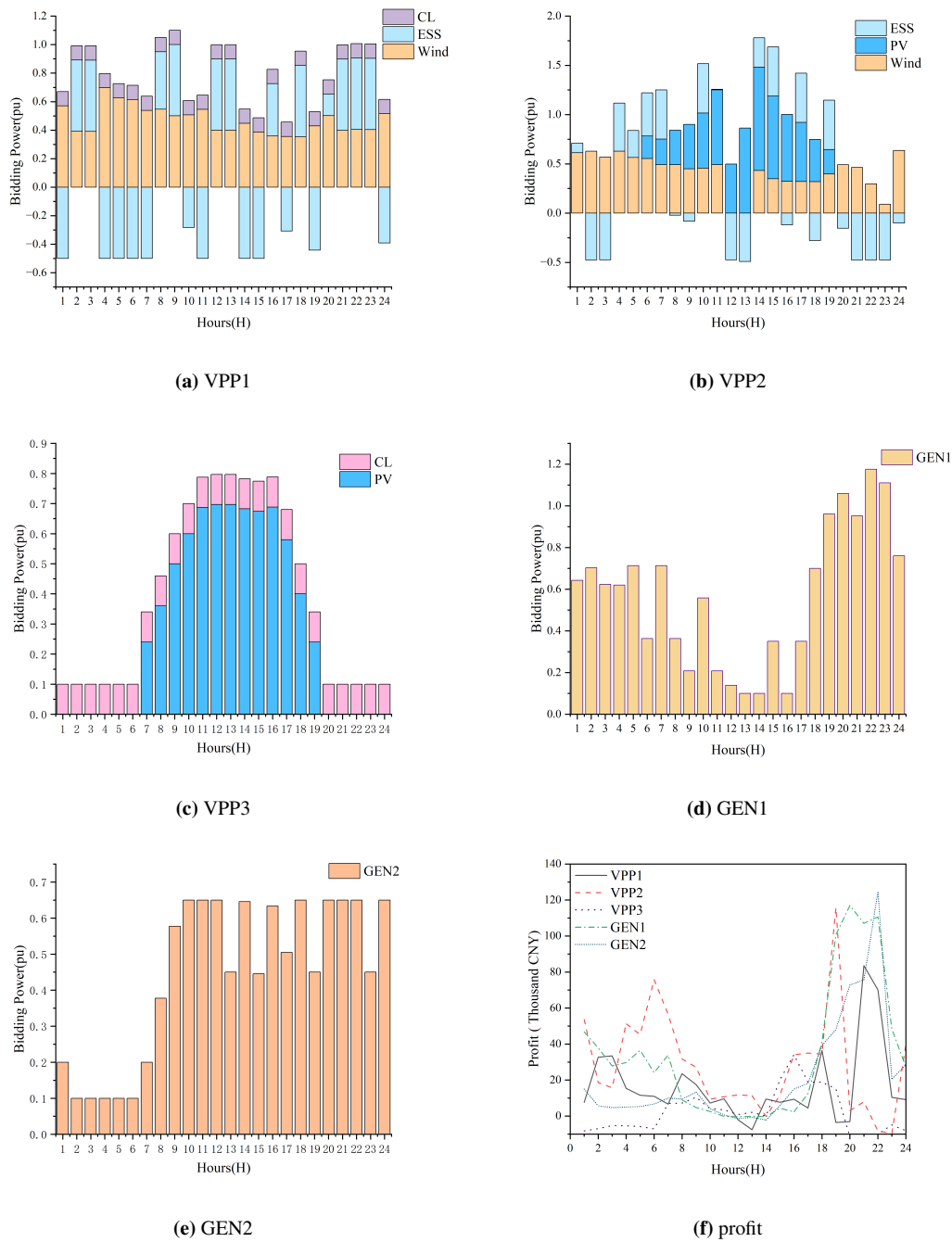
**FIGURE 6** Peak power and revenue of each unit

Fig. 6a–6e illustrate the 24-hour awarded status and awarded power of VPP1, VPP2, VPP3, thermal unit 1 (GEN1), and thermal unit 2 (GEN2), respectively. The awarded power shown in the figures is expressed in per-unit values with a base of 100 MW. Overall, the results indicate competition between renewable energy resources and thermal units in the energy market. Owing to their cost advantages, renewable energy resources exhibit a slight competitive edge. The controllable load power appears as a straight line because ancillary service markets are not considered in this study; at the current stage, China's electricity market is dominated by the energy market, and controllable loads are operated at their minimum power level to reduce costs.

The awarded power trajectories of thermal units 1 and 2 differ because GEN1 is located at a non-congested node and thus

does not experience congestion pricing; its awarded power drops sharply during periods of high photovoltaic output. In contrast, GEN2 is located at a severely congested node with only a single 65 MW transmission line to the grid. As a result, GEN2 bids at the minimum price to ensure clearing and profits from nodal congestion prices. Fig. 6f presents the revenues of all market participants in the energy market. The results show that unit revenues are generally positively correlated with nodal electricity prices, increasing during high-price periods. In this case study, the total outputs of renewable and conventional units are comparable, reflecting the trend that renewable installed capacity has increased year by year and exceeded that of thermal units. Consequently, thermal units lack a cost advantage in the energy market. In particular, VPP1 and VPP2, which aggregate more resources, achieve total daily revenues of CNY 401.3 thousand and CNY 683.2 thousand, respectively, while thermal units 1 and 2 earn CNY 824.2 thousand and CNY 524.5 thousand per day. Overall, the revenue gap between VPPs and thermal units in the energy market is not significant.

5 | CONCLUSIONS

This paper investigates the bidding behavior of multiple electricity market participants in the energy market and develops a modeling and solution framework based on the RM-MATD3 multi-agent reinforcement learning method. The study covers the construction of market game models under incomplete information and the design of reinforcement learning algorithms. Case studies are conducted to validate the effectiveness of the proposed approach and to further analyze its impacts on market clearing outcomes and strategic interactions. The main conclusions are summarized as follows:

(1) Under an incomplete-information electricity market setting, a bidding game model that captures multi-agent competitive behavior is constructed. Four representative multi-agent reinforcement learning algorithms—MATD3, MADDPG, MADDQN, and MADQN—are compared. The results demonstrate that MATD3, operating in continuous state and action spaces, outperforms the other three algorithms in terms of training stability and convergence speed. It exhibits stronger policy optimization capability and more accurately approximates Nash equilibrium solutions. Moreover, the delayed update scheme and the minimum Q-value mechanism in MATD3 effectively mitigate policy overestimation, enabling the model to maintain relatively high bidding profits in complex game environments.

(2) By incorporating real electricity price guidance into market bidding behavior, case studies confirm that the proposed A-MATD3 method better reflects actual market clearing conditions and allows market participants to engage in strategic competition within a reasonable range. Due to the existence of strategic interaction space, the dispersion of multi-agent actions is not strictly positively correlated with the deviation between simulated and real clearing prices. The value of the regularization coefficient significantly affects the alignment between simulated and real prices. The results indicate that selecting the coefficient within the range of 1–10 preserves sufficient strategic flexibility while avoiding excessive deviation from real prices, thereby enhancing the practical reference value of the simulation results.

6 | REFERENCES

REFERENCES

- [1] De K, Badar AQ. Virtual power plant profit maximization in day ahead market using different evolutionary optimization techniques. In: 2022 4th international conference on energy, power and environment (ICEPE) IEEE; 2022. p. 1–6.
- [2] Wang Y, Ai X, Tan Z, Yan L, Liu S. Interactive dispatch modes and bidding strategy of multiple virtual power plants based on demand response and game theory. *IEEE Transactions on Smart Grid* 2015;7(1):510–519.
- [3] Bakirtzis AG, Ziogos NP, Tellidou AC, Bakirtzis GA. Electricity producer offering strategies in day-ahead energy market with step-wise offers. *IEEE Transactions on Power Systems* 2007;22(4):1804–1818.
- [4] Kardakos EG, Simoglou CK, Bakirtzis AG. Optimal offering strategy of a virtual power plant: A stochastic bi-level approach. *IEEE Transactions on Smart Grid* 2015;7(2):794–806.
- [5] Chen R, Chen B. Research on the joint clearing model of electricity energy market and AGC market based on Stackelberg game. In: 2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI) IEEE; 2022. p. 139–142.
- [6] Rahimiyan M, Mashhadi HR. Supplier's optimal bidding strategy in electricity pay-as-bid auction: Comparison of the Q-learning and a model-based approach. *Electric Power Systems Research* 2008;78(1):165–175.

- [7] Liang Y, Guo C, Ding Z, Hua H. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm. *IEEE transactions on power systems* 2020;35(6):4180–4192.
- [8] Liu D, Gao Y, Wang W, Dong Z. Research on bidding strategy of thermal power companies in electricity market based on multi-agent deep deterministic policy gradient. *IEEE access* 2021;9:81750–81764.
- [9] Ye Y, Qiu D, Sun M, Papadaskalopoulos D, Strbac G. Deep reinforcement learning for strategic bidding in electricity markets. *IEEE Transactions on Smart Grid* 2019;11(2):1343–1355.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to all those who contributed to the completion of this work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Zeyu Li: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing—original draft preparation and writing—review and editing; Junhua Gu: resources, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

DATA AVAILABILITY

Data available on request from the authors.

REFERENCES

- [1] De K, Badar AQ. Virtual power plant profit maximization in day ahead market using different evolutionary optimization techniques. In: 2022 4th international conference on energy, power and environment (ICEPE) IEEE; 2022. p. 1–6.
- [2] Wang Y, Ai X, Tan Z, Yan L, Liu S. Interactive dispatch modes and bidding strategy of multiple virtual power plants based on demand response and game theory. *IEEE Transactions on Smart Grid* 2015;7(1):510–519.
- [3] Bakirtzis AG, Ziogos NP, Tellidou AC, Bakirtzis GA. Electricity producer offering strategies in day-ahead energy market with step-wise offers. *IEEE Transactions on Power Systems* 2007;22(4):1804–1818.
- [4] Kardakos EG, Simoglou CK, Bakirtzis AG. Optimal offering strategy of a virtual power plant: A stochastic bi-level approach. *IEEE Transactions on Smart Grid* 2015;7(2):794–806.
- [5] Chen R, Chen B. Research on the joint clearing model of electricity energy market and AGC market based on Stackelberg game. In: 2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI) IEEE; 2022. p. 139–142.
- [6] Rahimiyan M, Mashhadi HR. Supplier's optimal bidding strategy in electricity pay-as-bid auction: Comparison of the Q-learning and a model-based approach. *Electric Power Systems Research* 2008;78(1):165–175.
- [7] Liang Y, Guo C, Ding Z, Hua H. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm. *IEEE transactions on power systems* 2020;35(6):4180–4192.
- [8] Liu D, Gao Y, Wang W, Dong Z. Research on bidding strategy of thermal power companies in electricity market based on multi-agent deep deterministic policy gradient. *IEEE access* 2021;9:81750–81764.
- [9] Ye Y, Qiu D, Sun M, Papadaskalopoulos D, Strbac G. Deep reinforcement learning for strategic bidding in electricity markets. *IEEE Transactions on Smart Grid* 2019;11(2):1343–1355.



Junhua Gu is a doctoral supervisor at Hebei University of Technology, Tianjin, China. Her main research interests include data mining, image acquisition, reinforcement learning, and intelligent control.



Zeyu Li is currently pursuing the master's degree at Hebei University of Technology, Tianjin, China. His main research interests include reinforcement learning, electricity markets, and virtual power plants.