Continuous Active Power Dispatch Optimization for Virtual Power Plants with Energy Storage Using DDPG

Wenteng Liang[1], Yuheng Liu[1], Yizhen You[1], Sheng Xu[2]
[1]State Grid Jiangsu Electric Co. Ltd., Nanjing 211000, Jiangsu, China
[2]NARI-TECH Nanjing Control Systems Ltd., Nanjing 211106, Jiangsu, China

Corresponding author: Sheng Xu (e-mail: open202407@163.com)

ABSTRACT Continuous active power dispatch is a critical task for virtual power plants (VPPs) with energy storage, particularly under the increasing uncertainty of renewable generation and the growing demand for real-time coordination in smart energy systems. This study proposes an improved Deep Deterministic Policy Gradient (DDPG)-based optimization framework to enhance continuous dispatch performance for VPPs integrating wind power, photovoltaic generation, and battery energy storage. A multi-objective scheduling model is established by jointly considering dispatch cost, renewable energy utilization, and operational constraints, while priority experience replay, cross-attention mechanisms, gradient clipping, and cosine learning-rate decay are incorporated to improve training efficiency and policy stability. Simulation results demonstrate that the proposed approach significantly outperforms conventional methods by reducing average daily dispatch costs by 16.6%, achieving a renewable energy utilization rate of 96.8%, and maintaining rapid power balance recovery under highly volatile operating conditions. The optimized dispatch strategy further improves system robustness and dynamic adaptability in scenarios involving renewable fluctuations and sudden load variations. Beyond intelligent energy management, the proposed framework provides a practical optimization methodology for communication-enabled smart grids and distributed electromagnetic sensing infrastructures, where reliable information exchange and adaptive control are essential for coordinated operation of energy and wireless network resources.

INDEX TERMS virtual power plant, deep deterministic policy gradient, energy storage, active power dispatch, smart grid communication

I. INTRODUCTION

A. RESEARCH BACKGROUND

DEEP Reinforcement Learning (DRL) algorithms offer novel solutions for complex system optimization problems through their exceptional decision-making capabilities in high-dimensional continuous state and action spaces. This capacity for handling complexity is crucial for the energy management of the industry, enabling the accurate, continuous optimization required for balancing the dynamic load, variable generation, and storage assets in its high-dimensional VPP operational environment. Among these, the DDPG algorithm, as a representative Actor-Critic framework, directly outputs deterministic dispatch strategies without requiring discretization of continuous action spaces. This aligns more closely with the practical requirement for “continuous power regulation” in active power dispatch for virtual power plants [1-3]. Among these, the DDPG algorithm, as a representative Actor-Critic framework, directly outputs deterministic dispatch strategies without requiring discretization of continuous action spaces. This aligns more closely with the practical requirement for "continuous power regulation" in active power dispatch for virtual power plants. While existing research has applied DDPG to microgrid scheduling and energy storage optimization control, its application to continuous active power scheduling for virtual power plants incorporating energy storage exhibits several limitations: First, it inadequately accounts for the dynamic uncertainty of renewable energy output, often relying on static scenario assumptions that compromise robustness under fluctuating realworld conditions; Second, the charging/discharging losses and lifetime degradation models for energy storage systems are overly simplified, making it difficult to reflect the actual impact of scheduling strategies on storage economics. Third, issues such as low sample utilization and slow convergence during algorithm training limit their engineering application efficiency [4,5].

The theoretical significance of this research lies in establishing a comprehensive technical framework— “uncertainty modeling-multi-objective optimization-enhanced DDPG solution”—which fills the application gap of deep reinforcement learning in continuous active power scheduling for virtual power plants incorporating energy storage, thereby enriching the theoretical system of intelligent VPP scheduling. Its practical value lies in enabling the proposed improved DDPG algorithm to achieve dynamic collaborative optimization of generation, storage, and load within virtual power plants. This significantly enhances dispatch economic efficiency and environmental performance, providing technical support for virtual power plants to participate in electricity market transactions and deliver ancillary services. Simultaneously, it offers reference for optimizing power system dispatch under high renewable energy penetration. This contribution is vital for industrial decarbonization, specifically by enabling large, flexible consumers in the industry to participate more effectively in system dispatch, thereby stabilizing the grid and accelerating the shift toward renewable-powered manufacturing [6].

Among these, the DDPG algorithm, as a representative Actor-Critic framework, directly outputs deterministic dispatch strategies without requiring discretization of continuous action spaces. This aligns more closely with the practical requirement for "continuous power regulation" in active power dispatch for virtual power plants [1-3]. Specifically, Ref. [1] verifies the DDPG algorithm's inherent capability to handle continuous action spaces; Refs. [2,3] validate the applicability of this capability in VPP active power dispatch scenarios, forming a complete technical logic chain.

B. CURRENT RESEARCH STATUS AT HOME AND ABROAD

Internationally, the EU's "FENIX" project first proposed a three-tier architecture for virtual power plants—aggregation, dispatch, and trading—incorporating energy storage systems as key balancing resources into the active power dispatch system. It established an optimization model targeting dispatch cost minimization but did not account for the uncertainty of renewable energy output [7]. The Lawrence Berkeley National Laboratory in the United States proposed a robust dispatch model for virtual power plants incorporating energy storage, based on scenario analysis. This approach generates multiple renewable energy output scenarios to cover uncertainties, but the excessive number of scenarios leads to exponential growth in computational complexity [8]. In academia, Khalid et al. studied microgrid-based virtual power plants, constructing a dispatch model incorporating wind, solar, and storage. They employed opportunity constraint programming to address output uncertainty but achieved only single-objective optimization without considering environmental benefits [9].

Domestically, a Tsinghua University team proposed a "source-storage-load coordination" VPP scheduling framework. This dynamically links energy storage charging/discharging strategies with renewable output prediction errors to establish a multi-objective optimization function. However, it converts multiple objectives into a single objective using a weighted sum method, which suffers from strong subjectivity in weight assignment [10]. The State Grid Electric Power Research Institute developed an active power dispatch model for industrial park virtual power plants, incorporating grid peak-valley pricing to optimize energy storage timing and reduce electricity procurement costs. However, it did not address strategies for enhancing renewable energy integration rates [11]. A team from Jiangnan University developed a rolling dispatch model for virtual power plants based on Model Predictive Control (MPC). While capable of real-time prediction updates, it relies on traditional quadratic programming algorithms for continuous variable optimization, making it prone to local optima [12].

C. RESEARCH CONTENT AND TECHNICAL APPROACH

1) Research Content

Modeling and uncertainty analysis of energy storage-integrated VPP systems: Develop mathematical models for wind turbines, PV arrays, energy storage systems, and conventional loads. Combine probability distributions with real-time forecasting errors to characterize renewable output uncertainty. Define constraints for energy storage charging/discharging, grid interaction power exchange, and load supplydemand balance.

Multi-objective scheduling optimization function construction: Establish a dual-objective function minimizing the virtual power plant's average daily scheduling cost (including electricity procurement and energy storage loss costs) while maximizing renewable energy utilization rates. Resolve objective conflicts using nondomination sorting to generate a reward function directly trainable by DDPG algorithms;

2) Technical Approach

Modeling Phase (1-2 months): Define VPP system composition → Establish mathematical models for wind, solar, storage, and load → Quantify renewable energy output uncertainty → Construct multi-objective optimization function;

Algorithm Improvement Phase (2-3 months): Build standard DDPG algorithm framework → Introduce priority-based experience replay and attention mechanism → Add gradient clipping and learning rate decay strategies → Design reward function and state/action spaces;

Training and Simulation Phase (2 months): Collect historical VPP operational data (wind/solar output, load, electricity prices) → Partition into training and test sets → Train algorithm and tune parameters → Build dispatch simulation platform using MATLAB/Simulink;

Validation and Analysis Phase (1 month): Conduct dispatch simulation experiments under various scenarios → Compare performance metrics between improved DDPG and traditional algorithms → Analyze algorithm robustness and economic efficiency → Formulate final conclusions and optimization recommendations.

II. MODELING AND DISPATCH CONSTRAINTS FOR VIRTUAL POWER PLANT SYSTEMS WITH ENERGY STORAGE

A. VIRTUAL POWER PLANT SYSTEM COMPOSITION AND MATHEMATICAL MODEL

As the core carrier aggregating distributed energy resources, energy storage systems, and controllable loads, the system model of a virtual power plant with energy storage must accurately reflect the operational characteristics of each component unit to provide a reliable foundation for subsequent dispatch optimization [13]. The virtual power plant system constructed in this study comprises wind turbines, photovoltaic arrays, lithium-ion battery energy storage systems, and conventional loads within an industrial park. The mathematical models for each unit are designed as follows:

The output level of wind turbines is entirely dependent on wind speed variations. In actual operation, the probabilistic distribution characteristics of wind speed and the nonlinear relationship with output must be considered [14].

Furthermore, due to limitations in meteorological forecasting accuracy, deviations exist between actual and predicted wind power output. Historical data indicates these deviations typically range between -10% and +10%, following a normal distribution pattern. Therefore, a forecast error coefficient must be incorporated into the model to adjust the theoretical output of wind turbines to actual output, ensuring consistency between the model and real-world scenarios. Validation shows this model can control simulation errors for wind power output within 5%, meeting the accuracy requirements for dispatch optimization [15-17]. The output of photovoltaic arrays is influenced by both solar irradiance and ambient temperature, with solar irradiance having a more significant impact. The probability distribution of local solar irradiance exhibits a "high in the middle, low at both ends" characteristic, with irradiance typically ranging between 200 and 1000 watts per square meter during most periods. Peak irradiance reaches its maximum value of 1000 watts per square meter at noon. Under rated irradiance (1000 W/m^2) and rated temperature (25°C), the rated output of a PV array is 150 kW. When irradiance falls below 200 W/m^2, PV output drops significantly, maintaining only 10%-20% of rated output [18]. The summary of VPP Dispatch and Operation Constraints is shown in Table 1:

Environmental temperature negatively correlates with PV output: each 1°C increase in temperature reduces output by approximately 0.45%. For instance, when ambient temperatures reach 35°C during summer afternoons, actual PV output drops by about 4.5% compared to the rated value. while in low-temperature winter conditions, output approaches the rated value more closely. Similar to wind power, photovoltaic output also exhibits prediction errors, which are more significant due to factors like cloud cover and sudden weather changes, typically ranging from -15% to +15%. Therefore, corresponding prediction error coefficients must be incorporated into the model to correct theoretical photovoltaic output and ensure the authenticity of simulation results. Compared with actual operational data, this model achieves 94% accuracy in simulating output under sunny conditions and 88% under cloudy conditions, effectively supporting dispatch optimization analysis [19].

B. CONSTRUCTION OF MULTI-OBJECTIVE DISPATCH OPTIMIZATION FUNCTION

The core objective of active power scheduling for virtual power plants is to achieve synergistic optimization between economic efficiency and environmental performance. This requires simultaneously minimizing scheduling costs while maximizing renewable energy utilization rates. The specific function is constructed as follows:

Scheduling costs comprise two components: grid power purchase costs and energy storage charging/discharging loss costs. Power purchase costs are directly related to purchased power capacity, time-of-use electricity rates, and time step intervals. The local peak-valley time-of-use pricing policy is implemented as follows: Peak periods (9:00-12:00, 17:00-21:00): 1.2 RMB/kWh Off-peak periods (7:00-9:00, 12:00-17:00, 21:00-23:00): 0.8 RMB/kWh Off-peak (23:00-7:00): 0.4 yuan/kWh. For example, purchasing 100 kWh during peak hours incurs a 5-minute cost of 10 yuan (100 kWh × 1.2 yuan/kWh × 5/60 hours) [20]. Energy storage charging/discharging loss costs correlate with power levels, efficiency, and loss cost coefficients. Charging energy loss equals charging power multiplied by (1 - charging efficiency). Discharging energy loss equals discharging power multiplied by (1 / discharging efficiency - 1), then multiplied by the loss cost coefficient of 0.3 yuan/kWh and the time interval to determine the loss cost for that period. For example, charging at 100 kW for 5 minutes incurs a loss cost of approximately 0.46 yuan (100 kW × (1 - 0.92) × 0.3 yuan/kWh × 5/60 hours). Summing the electricity procurement costs and energy storage loss costs across all 288 scheduling periods (5 minutes per period) yields the average daily total scheduling cost, which must be minimized through scheduling optimization [21].

III. DESIGN OF ACTIVE POWER DISPATCH OPTIMIZATION ALGORITHM BASED ON IMPROVED DDPG

A. PRINCIPLES OF STANDARD DDPG ALGORITHM

The DDPG (Deep Deterministic Policy Gradient) algorithm, based on the Actor-Critic framework, is suitable for reinforcement learning problems with continuous action spaces. Its core architecture comprises four components: Actor network, Critic network, target Actor network, and target Critic network. Their functions and collaborative mechanisms are as follows:

The core function of the Actor network is to output deterministic actions based on the current environmental state, namely the energy storage charging/discharging power adjustment and grid interaction power adjustment. The network employs a three-layer fully connected neural network structure: The input layer consists of an 8-

TABLE 1. Summary of VPP Dispatch and Operation Constraints

Subcategory	Detailed Description
Overview	To ensure the safety, stability, and economic efficiency of VPP active power dispatch, various constraints must be defined to prevent dispatch strategies from violating equipment operation limits or grid control requirements.
Power Balance Constraint	As the core constraint of VPP dispatch, it requires real-time balance between active power supply and demand within the VPP, considering the interactive power among wind turbines, PV arrays, energy storage systems (ESS), conventional loads, and the main grid. Specifically, the sum of the actual output of wind and PV, power purchased from the grid, and ESS discharge power must equal the sum of the actual conventional load demand and ESS charging power. Breaking this balance will cause system frequency fluctuations and affect power supply stability. For example, when the output of wind and PV surges, ESS charging power should be increased or power sold to the grid should be raised to maintain balance; when the load surges, ESS discharge power should be increased or power purchased from the grid should be augmented to fill the gap.
Grid Interactive Power Constraint	The interactive power between the VPP and the main grid must comply with grid dispatch requirements to avoid impacts on grid frequency and voltage. According to local grid control standards, the maximum power purchase of the VPP is set to 200 kW (approximately 1.2 times the maximum load) to prevent local grid overload caused by excessive power purchase; the minimum power sale is set to -150 kW (the negative sign indicates power sales, and the absolute value is approximately 1.0 times the maximum renewable energy output) to avoid voltage impacts on the grid due to excessive power sales. In actual dispatch, the interactive power with the grid must be strictly controlled within this range: if it exceeds the upper limit, power purchase should be reduced or power sale increased; if it falls below the lower limit, power sale should be reduced or power purchase increased.
ESS Operation Constraints	In addition to the aforementioned SOC and charge-discharge power constraints, the ESS must also meet the charge-discharge power change rate constraint and operational safety constraints. The charge-discharge power change rate constraint is specified in the model: the power adjustment range between adjacent dispatch time steps shall not exceed 50 kW. Operational safety constraints mainly include avoiding deep charge-discharge and preventing over-temperature operation. Among them, the deep charge-discharge constraint is indirectly achieved through the SOC range (20%-80%); for the over-temperature operation constraint, the battery temperature is monitored in real time, and when the temperature exceeds 45°C, the charge-discharge power is automatically reduced to 50% of the rated value to ensure equipment safety.
Renewable Energy Output Constraints	The actual output of wind turbines and PV arrays must be within their technical allowable ranges, not exceeding the rated output upper limit or being negative. Specifically, the actual output of wind turbines should be controlled between 0-200 kW, and that of PV arrays between 0-150 kW. If the theoretical output of renewable energy exceeds the rated value, the output must be reduced to the rated value through curtailment; if the theoretical output is negative (affected by prediction errors), it should be corrected to 0 to ensure the rationality of the model output.
Overview	To ensure the safety, stability, and economic efficiency of VPP active power dispatch, various constraints must be defined to prevent dispatch strategies from violating equipment operation limits or grid control requirements.

dimensional state vector (wind power output, solar power output, load, SOC, energy storage power, grid interaction power, electricity price, renewable energy forecast error). The hidden layers contain 256 and 128 nodes respectively, utilizing the ReLU activation function. The output layer consists of a 2- dimensional action vector (energy storage power adjustment and grid interaction power adjustment), employing the tanh activation function to map action values to the interval [-1,1]. Scaling then determines the actual action range (energy storage power adjustment: [-50,50] kW; grid interaction power adjustment: [-60,60] kW) [22-24]. The Architecture and Technical Specifications of the Standard DDPG Algorithm are shown in Table 2:

The target Actor network shares identical architecture with the Actor network to generate target actions. Its parameters are not directly updated but replicated from the Actor network via a soft update strategy. The soft update coefficient is set to 0.001, meaning target Actor network parameters = 0.001 × Actor network parameters + 0.999 × previous target Actor network parameters. This strategy prevents excessive parameter fluctuations in the target network, ensuring training stability [25].

The core function of the Critic network is to evaluate the value of actions output by the Actor network, i.e., to compute action value based on the current state and action. The network also adopts a three-layer fully connected neural network structure: the input layer consists of a concatenated state-action vector (8- dimensional state + 2-dimensional action = 10 dimensions), the hidden layers have 256 and 128 nodes respectively, and the activation function is ReLU; the output layer is a single node that directly outputs the action value, reflecting the long-term reward of the current action in the current state [26].

The target Critic network shares the same architecture as the Critic network and is used to compute target action values. Its parameter updates follow the same soft update strategy as the target Actor network, with the soft update coefficient set to 0.001. Target action values are computed by combining the next-step state, the target action output by the target Actor network, and the current reward, thereby providing stable value targets for training the Critic network [26].

To break sample correlations and enhance training efficiency, the DDPG algorithm incorporates an experience replay mechanism: each scheduling time step's experience sample (current state, action, reward, next state, termination flag) is stored in an experience replay pool (capacity: 1 million). During training, batch samples (64 per batch) are randomly sampled from the replay pool for parameter updates. This mechanism prevents sample order from influencing training while enabling sample reuse, thereby improving data utilization efficiency [27].

TABLE 2. Architecture and Technical Specifications of the Standard DDPG Algorithm

Component of Standard DDPG Algorithm	Core Functionality	Technical Details and Implementation
Actor Network	Outputs deterministic actions based on current environmental states	• Network structure: 3-layer fully connected neural network (FCN). Input layer: state vector (dimension varies by task); hidden layers: 256/128 nodes with ReLU activation; output layer: action vector with tanh activation (maps actions to [-1,1] for subsequent scaling to actual action range). • Objective: Maximize the action value evaluated by the Critic network via policy gradient ascent.
Target Actor Network	Generates stable target actions for Critic network training	• Structure: Identical to the Actor network. • Parameter update: Uses soft update instead of direct update. Formula: $\theta_{\mu'} = \tau\theta_\mu + (1-\tau)\theta_{\mu'}$, where τ (typically 0.001) is the soft update coefficient, θ_μ is the Actor network parameter, and $\theta_{\mu'}$ is the Target Actor network parameter. This avoids large parameter fluctuations and ensures training stability.
Critic Network	Evaluates the value of actions output by the Actor network (action-value function $Q(s, a)$)	• Network structure: 3-layer FCN. Input layer: concatenated vector of state and action (dimension = state dimension + action dimension); hidden layers: 256/128 nodes with ReLU activation; output layer: single node (outputs the action value directly). • Objective: Minimize the mean squared error (MSE) between the predicted action value and the target action value.
Target Critic Network	Calculates stable target action values to guide Critic network training	• Structure: Identical to the Critic network. • Parameter update: Follows the same soft update rule as the Target Actor network: $\theta_{Q'} = \tau\theta_Q + (1-\tau)\theta_{Q'}$, where θ_Q is the Critic network parameter and $\theta_{Q'}$ is the Target Critic network parameter.
Experience Replay Buffer	Breaks sample correlation and improves data utilization efficiency	• Function: Stores experience samples $(s_t, a_t, R_t, s_{t+1}, d_t)$, where s_t = current state, a_t = action, R_t = reward, s_{t+1} = next state, and d_t = done flag (1 for terminal state, 0 otherwise). • Sampling strategy: Random sampling of mini-batches (typically 64–128 samples per batch) during training to avoid sequential sample bias. Common buffer capacity: 10^5–10^6 samples.
Loss Functions of Actor and Critic networks	Guide the update of networks	1. Critic Network Loss: MSE between predicted and target action values. Target action value formula: $y_i = R_i + \gamma(1-d_i)Q'(s_{i+1}, \mu'(s_{i+1}; \theta_{\mu'}); \theta_{Q'})$, where γ (typically 0.9–0.99) is the discount factor, Q' is the Target Critic net work output, and μ' is the Target Actor network output. Loss formula: yi=Ri+γ (1−di) Q′ (si+1,μ′ (si+1;θμ′);θQ′) 2. Actor Network Loss: Negative of the average action value (maximizes action value via gradient ascent). Loss formula: Lμ=−B1∑i=1BQ(si,μ(si;θμ);θQ)
Exploration Mechanism	Introduces randomness to action, avoids local optima, and explores the action space	Adds Gaussian noise (typically $N(0, 0.12)$ to the Actor network's output action during training: $a_{\text{explore}} = \mu(s; \theta_\mu) + \epsilon$, where ϵ is Gaussian noise. The noise intensity can be gradually reduced to balance exploration and exploitation in later training stages.

The core function of the Actor network is to output deterministic actions based on the current environmental state, namely the energy storage charging/discharging power adjustment and grid interaction power adjustment. The network employs a three-layer fully connected neural network structure: The input layer consists of an 8-dimensional state vector (wind power output, solar power output, load, SOC, energy storage power, grid interaction power, electricity price, renewable energy forecast error).

The core function of the Critic network is to evaluate the value of actions output by the Actor network, i.e., to compute action value based on the current state and action. The network also adopts a three-layer fully connected neural network structure: the input layer consists of a concatenated state-action vector (8- dimensional state + 2-dimensional action = 10 dimensions). Note: The state space itself is fixed at 8 dimensions; the 10-dimensional input of the Critic network is the concatenation result of state and action vectors.

B. IMPROVED DDPG ALGORITHM DESIGN

To address issues in standard DDPG algorithms for virtual power plant active power scheduling—including low sample utilization, insufficient capture of key features, and training instability—we implement enhancements across four dimensions: priority-based experience replay, attention mechanism integration, gradient clipping, and learning rate decay. Specific designs are as follows:

Additionally, a cosine annealing learning rate decay strategy is adopted to balance exploratory capability during early training with convergence accuracy in later stages. Initial learning rates are set to 0.0001 for the Actor network and 0.001 for the Critic network. As training iterations progress (maximum 100,000 iterations), the learning rate gradually decays along a cosine curve, ultimately reaching 10% of its initial value. For instance, after 50,000 iterations, the Actor network's learning rate drops to 0.000055 and the Critic network's to 0.00055. This strategy prevents parameter oscillations caused by excessively high learning rates early in training and avoids convergence difficulties due to overly high rates later on, ultimately improving the algorithm's final convergence accuracy by 10% [28].

To capture renewable energy prediction errors in real time, a Cross-Attention mechanism is integrated into the Critic network. The design details are as follows:

1) Input Feature Definition

Query vector (Q): Renewable energy prediction error sequence (dimension: T×3, T=number of time steps, 3=wind error + PV error + total error);

Key vector (K): Renewable energy output sequence (dimension: T×2, 2=wind output + PV output);

Value vector (V): Renewable energy output sequence (same as K).

Attention Weight Calculation: $\mathrm{Attention}(Q, K, V) = \mathrm{Softmax}\left(QK^T/\sqrt{d_k}\right)V$, where d_k is the dimension of K ($d_k = 2$), ensuring stable gradient propagation.

Integration Position: The Cross-Attention module is embedded in the input feature fusion stage of the Critic network, with the workflow: 1) Extract "renewable energy prediction error", "wind power output", and "PV output" from the 8-dimensional state vector; 2) Expand dimensions of the three features to form Q, K, V vectors; 3) Calculate attention weights to strengthen the contribution of error-related features; 4) Concatenate the weighted features with other state features (SOC, electricity price, etc.) to form the final input of the Critic network.

The attention mechanism enables the model to dynamically focus on time steps with large prediction errors, increasing the attention weight for such features by 30% compared to the standard model.

To align with virtual power plant (VPP) dispatch requirements, the state space and action space of the DDPG algorithm are explicitly defined to ensure algorithm outputs match real-world dispatch scenarios: The state space comprises eight normalized dimensions (mapped to the [-1,1] range), specifically: actual wind power output (0–200 kW), actual PV output (0–150 kW), actual conventional load (100–300 kW), energy storage SOC (20%–80%), current energy storage charge/discharge power (-100–100 kW, positive for charging, negative for discharging) virtual power plant interaction power with the grid (-150-200 kW, positive for power purchase, negative for power sale), current electricity price (0.4-1.2 CNY/kWh), and renewable energy output prediction error (0-15%) [29,30].

The action space comprises two continuous dimensions, also scaled as follows: energy storage charge/discharge power adjustment (-50 to 50 kW, negative values reduce charging/increase discharging, positive values increase charging/reduce discharging); grid interaction power adjustment (-60 to 60 kW, negative values reduce power purchase/increase power sale, positive values increase power purchase/reduce power sale). The dimensional design of the action space ensures scheduling strategy flexibility while preventing excessive system fluctuations caused by overly large action adjustments.

2) SOC Constraint Enforcement Mechanism

To ensure SOC operates within the safe range (20%-80%), a dual mechanism of “reward function penalty + action clipping” is implemented.

3) Reward Function Penalty

When SOC<20% or SOC>80%, a penalty term is triggered: $R_{\text{penalty}} = -50 \times |\text{SOC} - 50\%|$. This reduces the policy’s cumulative reward to discourage deep charge-discharge.

4) Action Clipping

After the Actor network outputs an action, a constraint check is added. If executing the action would cause SOC to exceed the threshold:

- When SOC$\leq$20%: Prohibit discharging actions, limit charging power increase to $\leq$20 kW;
- When SOC$\geq$80%: Prohibit charging actions, limit discharging power increase to $\leq$20 kW;

This ensures SOC remains within the safe range after action execution.

C. IMPROVED DDPG ALGORITHM TRAINING PROCESS

The training process for the improved DDPG algorithm must incorporate the temporal characteristics of the virtual power plant scheduling scenario to ensure efficient convergence and learning of optimal scheduling strategies. The specific steps are as follows:

1. Initialization Setup: Construct the Actor network, Critic network, target Actor network, and target Critic network. Initialize network parameters (target network parameters match the main network initially). Initialize the experience replay pool and set training hyperparameters (maximum iterations: 100,000; batch size: 64; discount factor: 0.95; soft update coefficient: 0.001, etc.).

2. Environment Interaction and Experience Collection: Treat one typical day (288 scheduling moments) as one training cycle. Initialize the VPP’s initial state (e.g., initial SOC=50%, initial output and load values set based on historical data). At each scheduling time step, the Actor network outputs an action based on the current state. A Gaussian noise term N(0, 0.12) is added to increase exploration, yielding an exploration action. This action is applied to the VPP system, updating state variables such as energy storage SOC and grid interaction power to obtain the next-step state. The reward function calculates the current-step reward, determining whether the cycle has ended (if yes, termination flag = 1; otherwise = 0). Store the experience sample (current state, exploration action, reward, next-step state, termination flag) into the experience replay pool.

3. Network Parameter Update: When the replay pool reaches 10,000 samples, initiate network update: Select 64 batch samples from the replay pool using priority sampling; Compute target action value (based on target Actor and target Critic networks); Compute the Critic network loss (mean squared error), minimize the loss using the Adam optimizer, update the Critic network parameters, and perform gradient clipping; Compute the policy gradient for the Actor network (based on the Critic network’s gradient for actions), maximize the action value using the Adam optimizer, update the Actor network parameters, and perform gradient clipping; Update the target Actor and target Critic network parameters according to the soft update policy.

IV. EXPERIMENTAL VALIDATION AND PERFORMANCE ANALYSIS

A. EXPERIMENTAL PLATFORM AND PARAMETER SETTINGS

The simulation platform was constructed using “MATLAB/Simulink + TensorFlow 2.8” to achieve collaborative virtual power plant system simulation and algorithm training: MATLAB/Simulink was employed to build system simulation models for wind turbines, photovoltaic arrays, energy storage systems, and conventional loads, performing power balance calculations, constraint condition evaluations, and state variable updates; TensorFlow 2.8 leveraged the Keras framework to build neural network models for the enhanced DDPG algorithm, enabling algorithm training and dispatch strategy output. Data exchange between the two systems was achieved through the MATLAB Engine API, facilitating real-time transmission of state variables and action variables to ensure synchronization between simulation and training.

The hardware configuration comprised: an Intel Core i9-12900K CPU (16 cores, 32 threads), NVIDIA RTX 4090 GPU (24GB VRAM), 64GB DDR5 memory, 2TB SSD storage. This configuration meets the efficiency demands of large-scale data processing and model training, with a single training cycle (288 scheduling moments) taking approximately 1.2 seconds and 100,000 iterations requiring roughly 33 hours total training time. Virtual Power Plant System Parameters Based on actual data from an industrial park, the parameters are as follows: Wind turbines: Rated output 200 kW, cut-in wind speed 3 m/s, rated wind speed 12 m/s, cut-out wind speed 25 m/s Photovoltaic arrays: Rated output 150 kW, temperature coefficient -0.0045/°C; energy storage system rated capacity 500 kWh, rated charge/discharge power 100 kW, charge/discharge efficiency 92%, SOC range 20%-80%; conventional load average daily peak 300 kW, valley 100 kW; grid interaction power constraint -150 to 200 kW, peak-valley time-of-use electricity price 0.4-1.2 RMB/kWh.

B. EXPERIMENTAL SCENARIO DESIGN

To comprehensively validate the performance of the improved DDPG algorithm, experimental scenarios were designed covering conventional operating conditions, extreme fluctuations, and sudden load events, The Experimental Environment and System Configuration for VPP Dispatch are shown in Table 3:

C. HIGHLY FLUCTUATING RENEWABLE ENERGY SCENARIO

Simulates extreme weather conditions (e.g., gusts, rapid cloud cover) to increase wind and solar power output prediction errors (wind error $\pm 15\%$, solar error $\pm 20\%$). Wind power output abruptly drops from 180 kW to 80 kW between 10:00–10:30, and PV output surged abruptly from 100 kW to 150 kW between 14:00 and 14:30. This scenario validated the algorithm's robustness (capability to handle renewable energy fluctuations).

D. LOAD SURGE SCENARIO

At 14:00 on a typical day, a 30% load surge (from 200 kW to 260 kW) was simulated to replicate sudden production demands in an industrial park, testing the algorithm's dynamic response capability (rapidly adjusting dispatch strategies to maintain power balance).

E. EXPERIMENTAL RESULTS AND ANALYSIS

First, we compare the convergence performance of the improved DDPG and standard DDPG algorithms, using the trend of average reward values during training as the evaluation metric (higher rewards indicate better scheduling performance). Experimental results show: The standard DDPG algorithm converged more slowly, requiring 80,000 iterations to stabilize, with a final reward value of only 7.2–7.8—over 15% lower than the improved DDPG algorithm.

Under highly fluctuating renewable energy conditions, the robustness of the four algorithms was analyzed using three evaluation metrics: "variation in wind and solar curtailment rates," "variation in dispatch costs," and "maximum power balance error" (where variation refers to the percentage difference between Scenario 2 and Scenario 1 metrics relative to Scenario 1 metrics). Results are as follows:

The modified DDPG algorithm demonstrated optimal robustness: the variation in wind and solar curtailment rates was only 1.8%, increasing by just 0.6 percentage points from 3.2% to 3.8% compared to Scenario 1; the variation in dispatch costs was 8.5%, increasing by 109.3 yuan from 1286.3 yuan to 1395.6 yuan; Maximum power balance error: 2.0%, remaining below the 3% safety threshold. This is attributed to: The improved DDPG algorithm employs an attention mechanism to capture renewable energy output prediction errors in real time, rapidly adjusting energy storage charging/discharging strategies—when wind power output drops sharply, it immediately increases energy storage discharge power (from 20 kW to 70 kW) and grid power purchase (from 30 kW to 80 kW) to maintain power balance; When solar output surged abruptly, it increased energy storage charging power (from 50 kW to 90 kW) and grid power sales (from 20 kW to 50 kW) to prevent curtailed generation.

Results confirm that the Cross-Attention mechanism significantly improves the model's ability to capture renewable energy prediction errors, enabling more timely adjustment of ESS charging/discharging strategies and further reducing scheduling costs.

In load surge scenarios, the dynamic response capabilities of four algorithms were analyzed using "power balance recovery time," "energy storage charge/discharge power adjustment range," and "peak grid power purchase" as evaluation metrics. Results are as follows:

The improved DDPG algorithm demonstrated the strongest dynamic response capability: The energy storage charging/discharging power adjustment range reached 45.3 kW, representing a 38.5% improvement over the standard DDPG (32.7 kW) and exceeding traditional algorithms by over 50%; The peak grid power purchase reached only 128.5 kW, a 17.8% reduction compared to the standard DDPG (156.3 kW) and over 30% lower than traditional algorithms.

Results show that the dual constraint mechanism effectively controls SOC within 20%-80% in 99.5%+ of scheduling time steps, with minimal out-of-bounds amplitude and duration, verifying compliance with physical constraints.

F. EXPERIMENTAL CONCLUSIONS

Based on results from three scenarios, the improved DDPG algorithm demonstrates significant advantages in continuous active power scheduling optimization for virtual power plants with energy storage:

1. Enhanced Economic Efficiency: Under conventional conditions, average daily scheduling costs are reduced by 16.6% compared to standard DDPG and by over 28% com-

TABLE 3. Experimental Environment and System Configuration for VPP Dispatch

Experimental Configuration Category	Specific Items	Detailed Specifications
Simulation Software Platform	Framework	Built on “MATLAB/Simulink + TensorFlow 2.8” to realize collaboration between VPP system simulation and algorithm training: • MATLAB/Simulink: Used to build system simulation models for wind turbines, PV arrays, ESS, and conventional loads; completes power balance calculation, constraint judgment, and state variable update. • TensorFlow 2.8 (based on Keras): Used to build the neural network model of the improved DDPG algorithm; realizes algorithm training and dispatch strategy output. • Data Interaction: MATLAB Engine API enables real-time data transmission of state variables and action variables between the two platforms, ensuring synchronization of simulation and training.
Environment Specifications	Hardware Component	• CPU: Intel Core i9-12900K (16 cores, 32 threads). • GPU: NVIDIA RTX 4090 (24 GB video memory). • RAM: 64 GB DDR5. • Hard Drive: 2 TB SSD. • Performance: Meets the efficiency requirements of large-scale data processing and model training; processing time for a single training cycle (288 dispatch time steps) is approximately 1.2 seconds, and total training time for 100,000 iterations is approximately 33 hours.
VPP System Parameters	Wind Turbine	Rated output: 200 kW; cut-in wind speed: 3 m/s; rated wind speed: 12 m/s; cut-out wind speed: 25 m/s.
VPP System Parameters	PV Array	Rated output: 150 kW; temperature coefficient: -0.0045/°C.
VPP System Parameters	Energy Storage System (ESS)	Rated capacity: 500 kWh; rated charge-discharge power: 100 kW; charge-discharge efficiency: 92%; SOC range: 20%-80%.
VPP System Parameters	Conventional Load	Daily average peak load: 300 kW; daily average valley load: 100 kW.
VPP System Parameters	Grid Interaction and Electricity Price	Grid interactive power constraint: -150 kW to 200 kW (negative value for power sales, positive value for power purchase); time-of-use electricity price: 0.4-1.2 CNY/kWh (valley, flat, peak periods respectively).

pared to traditional algorithms. This effectively leverages peak-valley electricity pricing and energy storage regulation to minimize electricity procurement expenses and energy storage losses.

2. Enhanced Environmental Performance: Achieves a renewable energy utilization rate of 96.8% with wind and solar curtailment rates as low as 3.2%, representing a 5-11 percentage point improvement over comparison algorithms, thereby supporting the achievement of carbon neutrality goals;

3. Superior Robustness and Dynamic Response: Under scenarios of high renewable energy volatility and sudden load surges, it achieves the shortest power balance recovery time (8.2 seconds), minimal power curtailment rates, and minimal cost fluctuations, ensuring stable operation under extreme conditions. This algorithm delivers an efficient, robust intelligent solution for continuous active power dispatch in virtual power plants incorporating energy storage. It can be further extended to dispatch scenarios involving virtual power plants with additional distributed energy types, providing technical support for the safe and stable operation of power systems and the integration of renewable energy.

V. CONCLUSION

This research successfully developed an Improved Deep Deterministic Policy Gradient (DDPG) algorithm for the continuous active power dispatch optimization of a Virtual Power Plant (VPP) integrating wind, PV, and Energy Storage Systems (ESS). This enhanced algorithm’s robust optimization is particularly effective when managing VPPs that include large industrial users, thereby enabling high-consumption complexes to leverage their demand flexibility and storage assets for maximized efficiency and grid stability. By systematically modeling VPP components, defining a multi-objective function (minimizing cost and maximizing renewable energy utilization), and applying a comprehensive set of operational constraints, this study established a robust framework for real-time dispatch. The core contribution is the enhancement of the standard DDPG through priority-based experience replay, an attention mechanism to capture renewable energy prediction errors, gradient clipping, and cosine annealing learning rate decay. Experimental validation across conventional, high-fluctuation, and load surge scenarios confirmed the superior performance of the improved algorithm. Compared to the standard DDPG and traditional methods, the enhanced DDPG significantly boosted economic efficiency, reducing average daily scheduling costs by 16.6%, primarily by optimally utilizing the time-of-use pricing and ESS regulation. Furthermore, it achieved an exceptional renewable energy utilization rate of 96.8%, demonstrating superior environmental performance. Critically, under extreme conditions, the improved algorithm showed superior robustness and dynamic response capability, achieving the shortest power balance recovery time (8.2 seconds) and minimizing grid power purchase peaks, ensuring system stability. These results validate the improved DDPG as an efficient, robust, and intelligent solution for continuous active power dispatch in VPPs, offering a promising foundation for the secure operation and high integration of distributed energy resources in future power systems. This validated solution is crucial for high-consumption manufacturing sectors, directly enabling the industry to securely integrate its onsite generation and flexible loads into VPPs, thereby optimizing its energy costs and contributing actively to grid stability.

AUTHOR CONTRIBUTIONS

Conceptualization – Wenteng Liang, Yuheng Liu, Yizhen You and Sheng Xu; methodology – Wenteng Liang, Yizhen You and Sheng Xu; investigation – Wenteng Liang, Yuheng Liu, Yizhen You and Sheng Xu; writing-original draft preparation – Wenteng Liang, Yuheng Liu, Yizhen You and Sheng Xu. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] M. Daneshmandi, M. Rahimiyan, and L. Baringo, "Medium-term energy scheduling of virtual power plants in futures markets considering energy and reserve pool participation," Energy Reports, vol. 15, Art. no. 109087, 2026, doi: 10.1016/J.EGYR.2026.109087.

[2] Z. Bao, S. Yu, and C. Tang, "Optimal dispatch strategy of virtual power plants with asymmetric information based on multi-agent deep reinforcement learning," Electric Power Systems Research, vol. 254, Art. no. 112580, 2026, doi: 10.1016/J.EPSR.2025.112580.

[3] L. Đorđe and Đ. Željko, "Stochastic scheduling of a sector-coupled virtual power plant integrating wind generation and thermal energy storage," Energy Conversion and Management: X, vol. 30, Art. no. 101638, 2026, doi: 10.1016/J.ECMX.2026.101638.

[4] W. Wang, S. Lin, B. Zhou, L. Qian, Y. Shen, and S. Muyeen, "Day-ahead coordinated operation strategy for a pricemaker virtual power plant with multiple prosumers," Electric Power Systems Research, vol. 255, Art. no. 112777, 2026, doi: 10.1016/J.EPSR.2026.112777.

[5] T. Gao, X. Wang, J. Zhang, J. Zhang, and A. Parisio, "Optimal scheduling of virtual power plant for short term operating reserve considering EV battery swapping," International Journal of Electrical Power and Energy Systems, vol. 175, Art. no. 111598, 2026, doi: 10.1016/J.IJEPES.2026.111598.

[6] Z. Peng and T. Heng, "Hierarchical optimization of virtual power plants via sequential Game-Based Multi-Agent reinforcement learning," Expert Systems With Applications, vol. 310, Art. no. 131263, 2026, doi: 10.1016/J.ESWA.2026.131263.

[7] V. Kumar, A. M. Niazi, U. Aslam, N. Saeed, M. Aurangzeb, and A. A. S. Syed, "A holistic optimization framework for virtual power plants with physics-informed battery degradation and probabilistic stability constraints," Future Batteries, vol. 9, Art. no. 100146, 2026, doi: 10.1016/J.FUB.2026.100146.

[8] R. Ma, T. Ma, Y. Hou, H. Luo, J. Liu, L. Li, et al., "Optimal Scheduling of a Park-Scale Virtual Power Plant Based on Thermoelectric Coupling and PV–EV Coordination," Eng, vol. 7, no. 1, pp. 54, 2026, doi: 10.3390/ENG7010054.

[9] Y. Dong, J. Su, Y. Xue, J. Zhao, S. Du, M. Dong, et al., "A Bidding Strategy to Address Risks of Virtual Power Plants in the Day-Ahead Electricity Market," International Transactions on Electrical Energy Systems, vol. 2026, no. 1, Art. no. 5283425, 2026, doi: 10.1155/ETEP/5283425.

[10] H. Zhang, D. Zhao, H. Pan, and L. Jia, "Optimal Bidding Strategy of Virtual Power Plant Incorporating Vehicle-to-Grid Electric Vehicles," Energies, vol. 19, no. 2, pp. 465, 2026, doi: 10.3390/EN19020465.

[11] X. Zheng, C. Ju, G. Yang, and J. Chu, "Risk-aware collaborative optimization for quantity–price bidding and energy storage dispatch in virtual power plants," Journal of Energy Storage, vol. 149, Art. no. 120112, 2026, doi: 10.1016/J.EST.2025.120112.

[12] X. S. Cao, L. Yue, G. Wang, G. C. Duan, K. Li, J. Li, et al., "A novel industrial load scheduling model to balance scheduling with virtual power plant regulation requirements," Sustainable Energy, Grids and Networks, vol. 45, Art. no. 102109, 2026, doi: 10.1016/J.SEGAN.2025.102109.

[13] X. Guo, Z. Chen, D. Ren, and Y. Liu, "Operation optimization of a virtual power plant integrated with electrical-thermal and electrical-hydrogen systems using nash bargaining: A multi-time scale approach," International Journal of Hydrogen Energy, vol. 205, Art. no. 153332, 2026, doi: 10.1016/J.IJHYDENE.2025.153332.

[14] M. Yang and T. Yu, "Uncertainty-driven IoT-based management of hydrogen-backed multi-carrier virtual power plants for zero-carbon microgrids," International Journal of Hydrogen Energy, vol. 205, Art. no. 153247, 2026, doi: 10.1016/J.IJHYDENE.2025.153247.

[15] T. Bowen, M. Blonsky, C. Simeone, R. McIlmoil, A. Alberg, and E. Estreich, "Design trade-offs for residential retail tariffs and virtual power plants," Applied Energy, vol. 407, Art. no. 127339, 2026, doi: 10.1016/J.APENERGY.2025.127339.

[16] L. Zhang, W. Guan, B. Wang, S. Wu, H. Wang, Y. Chang, et al., "Cost-optimal virtual power plant dispatch with hybrid storage strategy based on frequency-division coordination," Journal of Energy Storage, vol. 149, Art. no. 120302, 2026, doi: 10.1016/J.EST.2025.120302.

[17] L. Xie, G. Li, M. Xu, X. Dong, and Z. Tan, "Optimizing virtual power plant coordination through locational marginal flexibility under network constraints," Scientific reports, vol. 16, no. 1, pp. 3789, 2026, doi: 10.1038/S41598-025-33726-9.

[18] Y. Meng, H. Zhang, and X. Li, "Virtual power plant information-physical convergence network construction and its resilience assessment," Electrical Engineering, vol. 108, no. 2, pp. 78, 2026, doi: 10.1007/S00202-025-03417-9.

[19] F. Liu, Y. Xu, M. X. Zheng, and S. M. Wei, "Low-Carbon Scheduling Strategy for Multi-Virtual Power Plant Cluster Energy Sharing Collaborative Participation in Market Trading," Journal of Electrical Engineering & Technology, vol. (prepublish), pp. 1-25, 2026, doi: 10.1007/S42835-025-02562-2.

[20] S. Roy, C. D. Das, and N. Sinha, "MBGO-PID Based Frequency Regulation of Virtual Power Plants with Machine Learning Forecasted Day-Ahead Profiles for Diverse DER-Based Storage," Iranian Journal of Science and Technology, Transactions of Electrical Engineering, vol. (prepublish), pp. 1-35, 2026, doi: 10.1007/S40998-025-00981-5.

[21] S. Mei, X. Lu, and Q. Tan, "Dynamic aggregation optimization model of virtual power plant considering economic and environmental cost equilibrium," International Journal of Electrical Power and Energy Systems, vol. 174, Art. no. 111531, 2026, doi: 10.1016/J.IJEPES.2025.111531.

[22] P. Zhang and Y. Zheng, "Economical management of virtual power plant source-network-load-storage in the context of electricity-carbon market," PloS one, vol. 21, no. 1, Art. no. e0338321, 2026, doi: 10.1371/JOURNAL.PONE.0338321.

[23] M. S. Z. Khozaghi, A. Seifi, and B. G. Gharehpetian, "A new decentralized control structure for generation scheduling in a multi-energy virtual power plant considering its interactions with an electric vehicle aggregator and a demand response provider," International Journal of Electrical Power and Energy Systems, vol. 174, Art. no. 111539, 2026, doi: 10.1016/J.IJEPES.2025.111539.

[24] X. Ma, Z. Yumin, and Y. Jing, "Regulation and optimization of virtual power plants considering reversible energy storage of electricity and hydrogen under the new energy situation," Engineering Research Express, vol. 8, no. 2, Art. no. 025306, 2026, doi: 10.1088/2631-8695/AE321C.

[25] H. Wei and J. Zhang, "Contribution-driven cooperative trading strategy for multi-energy virtual power plants in the electricity-carbon coupled markets: An asymmetric Nash bargaining model," Energy Strategy Reviews, vol. 63, Art. no. 102047, 2026, doi: 10.1016/J.ESR.2026.102047.

[26] G. Xu, G. Yang, J. Bao, H. Feng, F. Zhang, and H. Zheng, "AI-enhanced multi-timescale optimization strategy for virtual power plants: Advancing losad forecasting and dynamic demand response integration," PloS one, vol. 21, no. 1, Art. no. e0339606, 2026, doi: 10.1371/JOURNAL.PONE.0339606.

[27] J. Jin, X. Pang, B. Wang, D. Wang, and Z. Zheng, "Optimal scheduling method of carbon-green certificate trading virtual power plant via Q-learning-enhanced particle swarm algorithm," Complex & Intelligent Systems, vol. 12, no. 1, pp. 51, 2025, doi: 10.1007/S40747-025-02176-1.

[28] B. Xie, L. Zhang, W. Zhao, Y. Yuan, X. Chen, X. Luo, et al., "A Study on Optimal Scheduling of Low-Carbon Virtual Power Plants Based on Dynamic Carbon Emission Factors," Sustainability, vol. 18, no. 1, pp. 326, 2025, doi: 10.3390/SU18010326.
[29] H. Qi, W. Sun, Y. Zhao, X. Qian, and X. Jiang, "Optimal Operation of Virtual Power Plants Based on Revenue Distribution and Risk Contribution," Energy Engineering, vol. 123, no. 1, pp. 1, 2025, doi: 10.32604/EE.2025.069603.
[30] Y. Li, Y. Chen, J. He, G. He, C. Dai, J. Tong, et al., "Optimal Dispatch of Urban Distribution Networks Considering Virtual Power Plant Coordination under Extreme Scenarios," Energy Engineering, vol. 123, no. 1, pp. 9, 2025, doi: 10.32604/EE.2025.069257.