Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 4, 2025

Adjusting the Dynamic Frequency Response in Photovoltaic-Energy Storage Grid-Connected Systems with the Help of Deep Deterministic Policy Gradient

Zhangyong Wei*

NR Electric Co., Ltd. Nanjing, 211102, Jiangsu, China
*Corresponding author's email: weizy0077@163.com

Abstract. This study applies the DDPG (Deep Deterministic Policy Gradient) algorithm to optimize the dynamic frequency response of PV (photovoltaic)-ES grid-connected systems under high-dimensional continuous control. A high-fidelity model incorporating PV output, ES SOC (State of Charge), and frequency dynamics was developed from operational data. An actor-critic architecture with target networks and experience replay trains the control strategy for precise ES charging/discharging. Compared to PID (Proportional-Integral-Derivative), fuzzy control, DQN (Deep Q-Network), and PPO (Proximal Policy Optimization), DDPG reduces maximum frequency deviation to 0.45 Hz, lowers average oscillation amplitude to 0.52 Hz, delivers a 7.1 s response, and achieves 85% charging/discharging efficiency and 75% ES utilization, extending SOC safe-range duration. DDPG enhances system stability and offers a cost-effective solution for new-energy frequency regulation.

Key words. Photovoltaic power generation, Energy-storage grid-connected systems, Dynamic frequency response, Frequency deviation, Deep deterministic policy gradient

1. Introduction

Against the backdrop of the global energy transition and the deepening of the low-carbon economy, the large-scale grid connection of renewable energy [1,2] has become an inevitable trend in the development of the power system. PV power generation [3,4] has been rapidly promoted due to its advantages of being clean, environmentally friendly and renewable, but its output power fluctuates significantly due to natural conditions such as weather and sunshine, posing a severe challenge to the stability of the power grid (PG) frequency. ES (energy storage) systems [5,6], as an important supplement to PV fluctuation regulation, can not only smooth PV output, but also achieve rapid frequency response, thereby ensuring the safe operation of the PG. Traditional frequency regulation methods often fail to meet the requirements due to insufficient response speed and regulation accuracy when facing high-frequency fluctuations and instantaneous load changes [7,8]. Therefore, exploring new intelligent control technologies has an important theoretical and practical significance.

In the current power system, PV power generation [9,10] has a random and volatile output, which leads to intensified grid frequency fluctuations, obvious frequency deviations and oscillations, and puts tremendous pressure on grid stability. Traditional PID control [11] or fuzzy control methods have problems such as insufficient response and inaccurate regulation when dealing with high dynamic changes, making it difficult to give full play to the frequency regulation advantages of ES systems. The various physical and economic constraints in the system, including ES power limitation, SOC management, grid-connected inverter capacity [12], and power slope, further increase the difficulty of regulation. Therefore, achieving fine regulation, rapid response, and efficient energy balance of the PV-ES grid-connected system under multiple constraints has become a key issue that needs to be solved urgently. As an important branch of deep reinforcement learning, the DDPG algorithm [13,14] provides a new idea for dynamic frequency response control with its excellent decision-making optimization ability in high-dimensional continuous action space. The DDPG algorithm can effectively suppress frequency deviation and oscillation under the premise of satisfying multiple physical and economic constraints, thus providing theoretical support and practical reference for the grid-connected control of new energy.

This paper applies the DDPG algorithm into the dynamic frequency response optimization of the PV-ES grid-connected system, and systematically constructs a frequency regulation control model that takes into account multiple constraints. In view of the volatility of PV output and the physical limitations of the ES system,

a comprehensive mathematical model including frequency regulation power balance, full response, ES power and SOC, inverter power limit and power slope constraints is constructed, which truly reflects the actual operation of the new energy grid-connected process. By designing an intelligent control strategy based on DDPG and through the careful construction of state, action and reward functions, the algorithm can converge quickly in a high-dimensional continuous control environment and achieve precise suppression of frequency deviation and oscillation. This paper uses the actual data of a province in northwest China as the background, uses high-frequency historical data and day-ahead forecast data to conduct simulation experiments, compares and analyzes traditional control methods and other reinforcement learning methods, and verifies the effectiveness and robustness of the proposed method in practical applications. The research results of this paper provide a new technical path for large-scale grid-connected frequency regulation of new energy, filling the research gap in the field of dynamic frequency regulation optimization under complex constraints.

This paper has a clear organizational structure and is divided into seven sections. Section 1: Introduction explains the frequency regulation challenges of photovoltaic-energy storage grid-connected systems and the shortcomings of traditional methods, and proposes the research value of the DDPG algorithm. Section 2: Related Works reviews the progress of traditional frequency regulation technologies and reinforcement learning, and points out the limitations of existing research. Section 3: PV-Energy Storage Grid-Connected System builds the system's physical architecture and mathematical model, covering photovoltaic output, energy storage SOC, inverter behavior, and grid frequency dynamic response. Section 4: DDPG Optimization details the design of the algorithm architecture, reward function, and constraint embedding mechanism, and explains the training process and strategy optimization method. Section 5: Example Simulation constructs a simulation environment based on real-world data to compare the performance metrics of DDPG and traditional methods. Section 6: Results and Analysis verifies the advantages of the algorithm from multiple dimensions, including frequency deviation, oscillation, response time, energy storage efficiency, and SOC management. Section 7: Conclusions summarizes the findings and outlines future directions for optimization. The full paper systematically demonstrates the efficiency and feasibility of DDPG in dynamic frequency regulation through theoretical modeling, algorithm innovation, and experimental verification.

2. Related Works

Traditional frequency regulation methods mainly rely on classical control technologies such as PID control [15], fuzzy control [16] and robust control, which can achieve basic frequency regulation under a single working condition. When faced with intermittent and volatile renewable energy such as PV power generation [17,18], traditional control strategies often fail to meet the instantaneous response requirements of the PG, and there are problems such as response delay and inaccurate regulation amplitude. To meet the above challenges, advanced methods such as model predictive control, optimization algorithm and adaptive control have been proposed to achieve precise control of frequency regulation while ensuring safe operation of the PG. In addition, as an important supplement to smoothing PV output and alleviating grid fluctuations, the role of ES systems [19,20] in dynamic frequency response has gradually been valued. By establishing a joint mathematical model of PVs, ES and loads, this paper explores how to achieve system power balance and frequency stability under multiple constraints. Through in-depth analysis of the ES charging and discharging process [21,22], SOC management, and grid-connected inverter characteristics, the influence of various constraints on frequency response performance is revealed. Although existing research has achieved certain results in theoretical modeling and control algorithms, most of them rely on traditional optimization methods, which makes it difficult to take into account response speed, regulation accuracy, and multiple operating constraints at the same time, limiting its application effect under complex working conditions. This provides a theoretical basis and research space for the further introduction of new intelligent control strategies, prompting the academic community to begin to focus on using deep learning and reinforcement learning methods to improve traditional frequency regulation strategies, and strive to achieve fast and precise dynamic regulation while meeting system safety and economic requirements.

With the rapid development of artificial intelligence and deep learning technologies, the application of deep reinforcement learning methods [23,24] in the field of power system control has gradually emerged and has become an important means to solve the shortcomings of traditional control methods. In particular, DDPG [25,26] has attracted more and more attention from scholars due to its outstanding performance in continuous action space. When dealing with high-dimensional, multi-constrained problems, DDPG [27,28] can directly output continuous control quantities by constructing an actor-critic network structure to achieve precise regulation of the charging and discharging process of the ES system, thereby reducing grid frequency deviation and oscillation. In terms of microgrid dispatching, energy management, and frequency control, the DDPG method [29,30] has shown significant advantages. Its rapid response and online learning capabilities provide an effective solution to the uncertainty brought about by the access of new energy. In view of the physical constraints and economic indicators of the system, scholars have gradually explored embedding multiple constraints into the reward function design to achieve real-time monitoring and optimization control of key parameters such as ES SOC, charging and discharging power, inverter capacity [31,32], and power slope [33,34]. Current research mainly focuses on optimization under a single control objective or simplified constraints, and has not fully

considered the comprehensive impact of various constraints in actual system operation. Therefore, under strict physical and economic constraints, using the DDPG algorithm to achieve high-precision, fast-response dynamic frequency regulation is still a difficult problem that needs to be solved.

3. PV-ES Grid-Connected System

A. PV-ES Grid-Connected Power Generation

The PV-ES grid-connected power generation system plays multiple roles in the power system. It not only promotes the high-proportion access of new energy, but also improves the frequency regulation capability and operational reliability of the system. Through the effective regulation of PV output fluctuations by the ES system, the system can quickly absorb or release electrical energy, smooth out grid frequency fluctuations, and ensure that the system remains stable under load changes and power generation fluctuations. In addition, its intelligent control mechanism enables efficient coordination of energy production, storage and consumption, optimizes the grid operation structure, and reduces the demand for backup capacity.

The PV-ES grid-connected power generation system is shown in Figure 1.

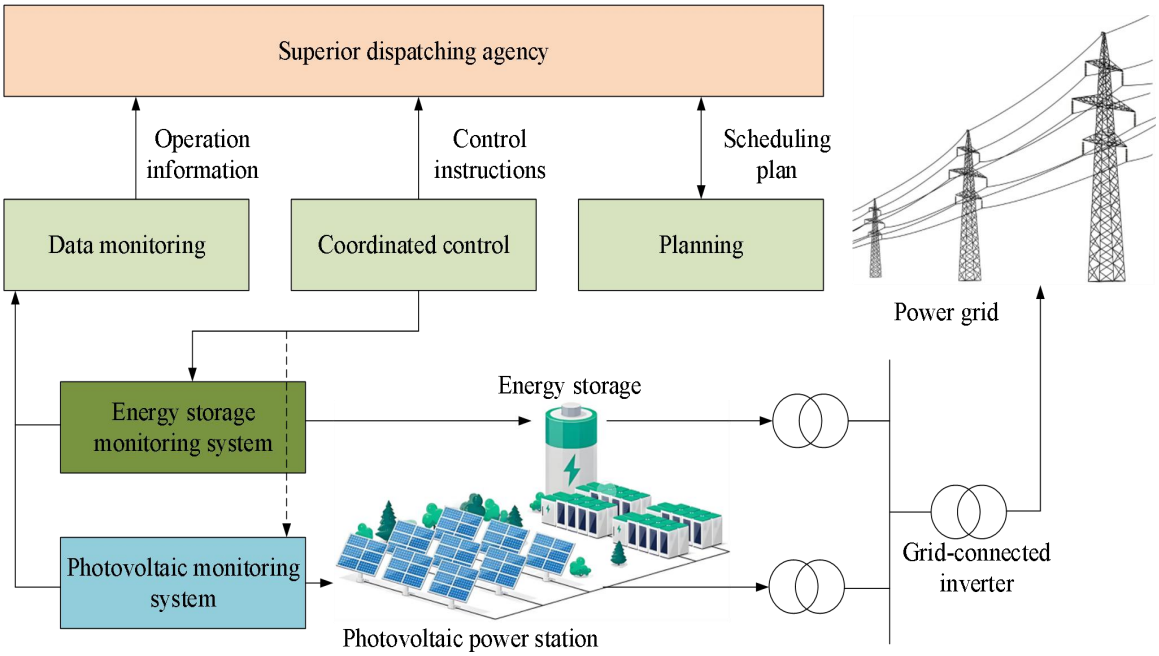

Figure 1. PV-ES grid-connected power generation system.

In the PV-ES grid-connected power generation system, the superior dispatching agency plays an important role in overall planning and decision-making, and is responsible for formulating the overall operation plan and emergency dispatching strategy. The data monitoring system can collect the operating status, environmental parameters and power output information of PV power stations and ES power stations in real time, providing accurate data support for subsequent coordinated control. The coordination control module analyzes and comprehensively judges the collected data in real time according to the preset plan code, coordinates the energy flow between the subsystems, adjusts the output of the grid-connected inverter through optimization instructions, and realizes efficient connection and energy exchange with the PG. PV power stations use advanced PV modules to convert solar energy into direct current, which is then connected to the PG after conversion by the inverter. ES power stations store and release electric energy through high-efficiency battery packs, balance load fluctuations, and ensure the stability of the grid frequency. The various components work together to form an intelligent and integrated energy management system, providing technical support and security for the high-proportion access of new energy.

B. Mathematical Model Establishment

The output power of PV modules is affected by environmental factors such as irradiance and temperature. Its basic output power model is described as:

$$P_{\mathrm{PV}}(t)=P_{\mathrm{PV,ref}}\cdot\frac{I(t)}{I_{\mathrm{ref}}}\cdot\left[1+\alpha\left(T(t)-T_{\mathrm{ref}}\right)\right] \quad (1)$$

$P_{\mathrm{PV}}(t)$ is the actual output power of the photovoltaic module at time t. $P_{\mathrm{PV,ref}}$ is the nominal output power under reference conditions. $I(t)$ is the actual solar irradiance at time t. I_{ref} is the reference irradiance. α is the temperature coefficient, which reflects the effect of temperature change on the output power. $T(t)$ is the actual temperature of the module at time t.

The mathematical model of the ES system is mainly used to describe the charging and discharging process and the evolution of the energy state. Its basic form is:

$$SOC(t+\Delta t)=SOC(t)+\frac{\eta_{ch}P_{ch}(t)-\frac{P_{dis}(t)}{\eta_{dis}}}{E_{rated}}\Delta t-D_{self}(SOC(t)) \quad (2)$$

The self-discharge characteristics are described by the function $D_{self}(SOC(t))$, considering the constant proportional decay:

$$D_{self}(SOC(t))=\gamma\cdot SOC(t) \quad (3)$$

In formula 3, γ is the self-discharge rate.

The inverter converts DC (direct current) power into AC (alternating current) power, and its conversion efficiency is expressed as:

$$P_{AC}(t)=\eta_{inv}\cdot P_{DC}(t) \quad (4)$$

In Formula 4, η_{inv} is the inverter efficiency.

$$P_{AC}(t)\le P_{inv,max} \quad (5)$$

By satisfying Formula 5, the inverter is prevented from overloading.

For the grid frequency response model, a differential equation based on power-frequency characteristics is used to describe it. The grid has an inertia constant M and a damping coefficient D, and the grid frequency dynamic response is:

$$M\frac{d\Delta f(t)}{dt}+D\Delta f(t)=\Delta P(t) \quad (6)$$

In formula 6, $\Delta f(t)$ is the frequency deviation, and $\Delta P(t)$ is the power regulation of the system at time t, which reflects the energy exchange and frequency regulation relationship between the inverter and the grid, and provides a dynamic response indicator for optimizing the control strategy.

4. DDPG Optimization

A. Algorithm Architecture and Reward Function

In the frequency regulation control of PV-ES systems, since the system state and control variables are continuous, traditional discrete action space reinforcement learning algorithms are difficult to apply directly. This paper adopts the DDPG algorithm, which can achieve efficient policy optimization in continuous action space.

The DDPG optimization process is shown in Figure 2.

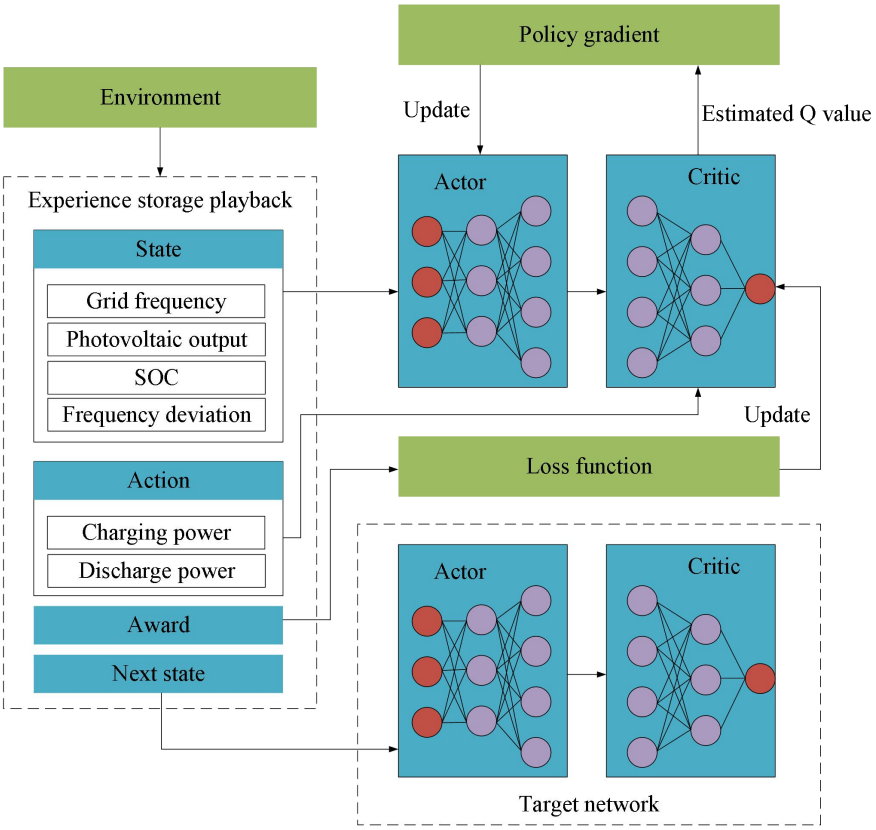

Figure 2. DDPG optimization process.

The target network in DDPG is a soft-updated delayed copy of the actor and critic. The target network is used in DDPG to generate the target Q value, which acts on the loss target of the critic. Through soft updates, it helps reduce the correlation and non-stationarity between the actor and the critic. The stabilizing effect of the target network is reflected in: by smoothing changes and decoupling the target from the rapidly changing online network, the divergence of the model is slowed down. This can greatly improve the stability and convergence of the model.

In the DDPG algorithm, the target network is a delayed copy of the online Actor and Critic networks, called the target Actor and the target Critic, respectively. They only gradually synchronize the online network parameters with a small soft update coefficient τ after each training, rather than completely copying them, so as to provide a relatively stable Q target value in the Critic loss calculation. Specifically, the target Critic calculates Q_target in combination with the action output of the target Actor in the next state, and optimizes the mean square error with the current Critic estimate. This delayed update mechanism effectively reduces the non-stationarity of the target value and reduces the training variance, thereby improving the convergence and stability of the algorithm.

In the DDPG algorithm architecture used in this study, the Actor network uses a three-layer fully connected layer structure (256-256-128 neurons), the hidden layer uses the ReLU activation function to enhance the nonlinear expression ability, and the output layer uses the Tanh function to map the action to [-1,1] and linearly scale it to the charge and discharge power range; the Critic network uses a dual-input branch structure (state input 256-128-64, action input 64-32), and after splicing, it outputs the Q value through two layers of fully connected layers (128-64), all using ReLU activation. The network is initialized using the He normal distribution, the optimizer uses Adam (Actor learning rate 0.0001, Critic learning rate 0.001), and the target network synchronizes parameters through soft updates (τ =0.001). Batch normalization is used to process state input during training, and Ornstein-Uhlenbeck noise (θ =0.15, σ =0.2) is added to the Actor output layer to improve exploration efficiency. The experience replay buffer capacity is set to 1e6, and the batch sampling size is 64, which fully implements the continuous control mechanism of the deep deterministic policy gradient algorithm.

The goal of the Critic network is to minimize the MSE (Mean Squared Error) loss function, and the update rule is:

$$L\left(\theta^{Q}\right)=\mathbb{E}_{(s,a,r,s)^{'}\sim D}\left[\left(Q\left(s,a\left|\theta^{Q}\right.\right)-y\right)^{2}\right] \quad (7)$$

The target value y is given by the Bellman equation:

$$y=r+\gamma Q^{'}\left(s^{'},\mu^{'}\left(s^{'}\left|\theta^{\mu^{'}}\right.\right)\right|\theta^{Q^{'}}\right) \quad (8)$$

In Formula 8, γ is a discount factor used to balance short-term and long-term rewards.

The Actor network optimizes the policy by maximizing the value estimated by the Critic network:

$$\nabla_{\theta^{\mu}}J=\mathbb{E}_{s\sim D}\left[\nabla_{a}Q\left(s,a\left|\theta^{Q}\right.\right)\right|_{a=\mu\left(s\left|\theta^{\mu}\right.\right)}\nabla_{\theta^{\mu}}\mu\left(s\left|\theta^{\mu}\right.\right)\Bigg] \quad (9)$$

The policy parameter θ^{μ} is updated through gradient ascent, so that the policy action a in state s can obtain a higher Q value.

To prevent instability during the strategy update process, DDPG uses a target network to stabilize training. The target network parameters are soft updated:

$$\theta^{Q^{'}}\leftarrow\tau\theta^{Q}+\left(1-\tau\right)\theta^{Q^{'}} \quad (10)$$

$$\theta^{\mu^{'}}\leftarrow\tau\theta^{\mu}+\left(1-\tau\right)\theta^{\mu^{'}} \quad (11)$$

In the PV-ES joint frequency regulation system, the key to reinforcement learning is to reasonably design the state, action and reward function so that the agent can learn an effective regulation strategy.

The system state needs to be able to fully reflect the dynamic characteristics of the PG, including: grid frequency, PV output, SOC of the ES system, and historical frequency deviation. The state variable is expressed as:

$$s_{t}=\left[f_{t},P_{\mathrm{PV}}\left(t\right),SOC_{t},\Delta f\left(t-1\right),\Delta f\left(t-2\right),\cdots\right] \quad (12)$$

The actions of the agent represent the charging and discharging power of the ES system, satisfying the physical constraints of the equipment:

$$P_{\min}\le a_{t}\le P_{\max} \quad (13)$$

In formula 13, $P_{\min}$ and $P_{\max}$ are the maximum discharge capacity and maximum charging power, respectively.

In the output layer of the Actor network, the Tanh activation function is used to normalize the action output to [-1,1] and map it to the charging and discharging power range through linear transformation:

$$a_t = P_{\min} + \frac{P_{\max} - P_{\min}}{2} \cdot \tanh\left(\mu\left(s_t \middle| \theta^{\mu}\right)\right) \quad (14)$$

The design of the reward function balances two aspects: reducing frequency deviation and reducing oscillation to ensure that the ES system operates within a safe range. The basic reward item can be designed as the negative sum of squared frequency deviations, with the formula:

$$r_t = -\left(\Delta f_t\right)^2 \quad (15)$$

$$\Delta f_t = f_t - f_{\text{ref}} \quad (16)$$

In formula 16, f_{ref} is the rated frequency.

To reduce oscillation, an oscillation amplitude penalty term is added:

$$r_t = -\left(\Delta f_t\right)^2 - \beta\left|\Delta f_t - \Delta f_{t-1}\right| \quad (17)$$

In Formula 17, β is a weight coefficient used to adjust the impact of the oscillation penalty term.

The final reward function is expressed as:

$$r_t = -\left(\Delta f_t\right)^2 - \beta\left|\Delta f_t - \Delta f_{t-1}\right| - \lambda \cdot \phi\left(s_t, a_t\right) \quad (18)$$

$\phi\left(s_t, a_t\right)$ is a constraint penalty term.

B. Constraints and Training Process

In the intelligent frequency modulation control based on DDPG, multiple constraints are comprehensively considered to ensure the response performance and operation safety of the system. The following constraints are embedded in the reward function so that the agent can take into account the steady-state and dynamic characteristics of the system during the optimization process.

Constraints on the balance of power involved in frequency modulation:

$$\sum_{i=1}^{N} \Delta P_i = \Delta P_{\text{demand}} \quad (19)$$

In formula 19, ΔP_i represents the power change of the i-th frequency modulation device. ΔP_{demand} is the power demand caused by load change.

Frequency modulation resources should be able to provide the required response capability within the specified time to ensure the frequency stability of the system:

$$T_{\text{response}} \le T_{\text{limit}} \quad (20)$$

In formula 20, T_{response} is the resource response time, and T_{limit} is the specified maximum allowable response time.

The charging and discharging power of the ES system is subject to the maximum power limit:

$$-P_{\max}^{\text{ch}} \le P_{\text{storage}} \le P_{\max}^{\text{dis}} \quad (21)$$

The state of charge of the ES system must be kept within a reasonable range:

$$SOC_{\min} \le SOC_t \le SOC_{\max} \quad (22)$$

The output power of the inverter is limited by the rated power, and the actual output power of the inverter is less than the rated power. In order to avoid excessive power mutations affecting system stability, the power change rate is limited:

$$\left|\frac{\mathrm{d}P}{\mathrm{d}t}\right| \le R_{\max} \quad (23)$$

In formula 23, $\frac{\mathrm{d}P}{\mathrm{d}t}$ is the power change rate, and $R_{\max}$ is the maximum allowable power change rate.

The ultimate goal of the frequency modulation system is to maintain the system frequency within the allowable range, and the frequency deviation should be minimized:

$$\left|\Delta f\right| \le f_{\text{limit}} \quad (24)$$

The power of the PV system is affected by weather and equipment conditions. Consider the maximum output power limit:

$$P_{\text{PV}} \le P_{\text{PV}}^{\max} \quad (25)$$

In order to improve the stability of training, an experience replay buffer is established to store historical interaction data to break the impact of data correlation on training and improve sampling efficiency.

In the model optimization phase, the training process randomly samples a batch of data from the experience replay buffer and uses the Critic network to calculate the value estimate of the current policy. In order to make the Critic network's value assessment of the environment more accurate, the training process continuously

minimizes the temporal difference error and adjusts the parameters of the Critic network to make its estimated value closer to the actual reward value.

In order to ensure the convergence and stability of the algorithm, the DDPG algorithm uses a target network mechanism. The target network is a copy of the Actor network and the Critic network, but its parameter update speed is slow and is updated using the exponential sliding average method. This approach is used to prevent instability caused by rapid changes in strategies during training and improve the training efficiency and convergence of the algorithm.

5. Example Simulation

A. Basic Data

This paper selects a PV-ES grid-connected system actually operated in a province in northwest China as the research object, and uses the actual PV output data and day-ahead forecast data of the region for 10 consecutive days to build a simulation environment. In the calculation example, the installed capacity of the PV power station is 90MW, and the real-time frequency data of the system is directly obtained from the website of the British Elexon company [35] to ensure the authority and timeliness of the data source. Through multi-source data fusion, the regional climate, irradiance fluctuations and the response characteristics of the ES system are fully considered, and the dynamic frequency regulation performance of the system under the condition of high proportion of PV access is deeply analyzed.

The data in this article are derived from the actual operation data of the photovoltaic-energy storage system in a province in northwest China and the public frequency data of the UK Elexon power grid. After cleaning, a simulation environment is constructed. The parameter settings are in line with engineering practice and physical constraints to ensure the non-sensitivity, legality and public availability of the data.

The calculation example parameters are shown in Table 1.

Table 1. Example parameters.

Parameter	Value	Parameters	Value
Data duration	10 days	Day-ahead forecast lead time	24 hours
Time interval	1min	ES discharge efficiency	80%
ES lifespan	18 years	Inverter efficiency	96%
ES charging efficiency	82%	ES maximum discharge power	45MW
Capacity cost	4805 yuan/MWh	ES minimum SOC	20%
Self discharge rate	0.03%	ES rated energy capacity	150MWh
PV installed capacity	90MW	Maximum power slope	0.5MW/min
Accuracy of irradiance measurement	±5%	Ambient operating temperature range	-10℃ ~ 50℃

The collected historical output data and frequency data are fully cleaned to ensure the quality and reliability of the input data. In view of the missing value problem in the original data, the mean filling method is used to reasonably fill it in to ensure data continuity. For the detected outliers, such as data points with sudden changes or excessive noise, Z-Score analysis is used to eliminate or correct them to avoid model deviation caused by abnormal data. The data from different data sources are converted into a unified format and time-aligned to eliminate the inconsistency between data collection and ensure that the simulation platform has high accuracy and reliability when simulating the actual system operation status.

In this study, mean filling was used to handle missing values in the data preprocessing stage, and abnormal data points were removed based on the Z-Score method (threshold ±3). Comparing the data distribution before and after cleaning, the original PV output power data showed a right-skewed distribution (skewness = 1.82), and the skewness was reduced to 1.05 after mean filling, but it introduced an underestimation of the tail fluctuation; Z-Score processing reduced the data standard deviation from 0.23MW to 0.18MW, which may have filtered out some effective extreme values.

Based on the pre-processed historical data and forecast data, a mathematical model of the PV module, ES system, and inverter and grid interface is established to fully reflect the dynamic characteristics and constraints of each component in actual operation. The above physical model is coupled with the model describing the frequency response characteristics of the PG to form a complete system simulation framework. The DDPG algorithm is embedded in this framework, and through real-time interaction with the simulation environment, the system status is collected, control actions are executed, and feedback rewards are obtained to achieve dynamic optimization of the frequency regulation process. Through this comprehensive simulation platform, the effectiveness of the algorithm in suppressing frequency deviation and oscillation in actual systems is effectively verified, providing a scientific basis and optimization strategy for the frequency regulation control of PV-ES grid-connected systems.

This study faces hardware limitations in edge deployment: the memory limit of edge devices (<8GB) prompts the use of model distillation technology to compress network parameters to 42% of the original size, and compresses the inference delay to 18.9ms through parameter quantization (8-bit fixed-point operation) and computational graph optimization.

This study implements explicit modeling of high-dimensional constraints in the DDPG framework by constraining embedded reward functions and dynamic weight adjustment mechanisms: the reward function integrates frequency deviation, oscillation penalty and SOC safety constraints (β =0.15, λ =0.5). Compared with the dual Q network of TD3 and the maximum entropy method of SAC, the violation rate of multi-dimensional physical/economic constraints is reduced while ensuring policy convergence. Through the soft update mechanism (τ =0.001) and constraint-aware sampling of experience replay, the constraint satisfaction rate and control accuracy in high-dimensional continuous action space are significantly improved.

B. Model Training and Evaluation

During the DDPG model training process, the Actor network and the Critic network are initialized to ensure that the model parameters have a good initial distribution. The initialization parameters include network structure, learning rate, discount factor, and soft update coefficient, which play a key role in the training effect. In actual operation, an experience replay buffer is established to store historical interaction data, and random sampling is used to break the time correlation between data. In the simulation environment, the preset exploration strategy is used to enable the agent to fully explore the continuous action space. To ensure the stability and convergence of the algorithm, the target network is used for soft update, so that the target network parameters slowly track the changes in the main network parameters to avoid drastic fluctuations during training.

DDPG initialization parameters are shown in Table 2.

Table 2. DDPG parameter initialization.

Parameter name	Parameter value	Function
Actor learning rate	0.0001	Control the update speed of actor network parameters, affecting strategy output
Critic learning rate	0.001	Control the update speed of critic network parameters, affecting value estimation
Discount factor	0.99	Balance short-term and long-term rewards, determine future reward discounts
Soft update factor	0.001	Ensure smooth update of target network and stabilize training process
Batch size	64	The amount of data sampled during each update affects update efficiency
Replay buffer size	1000000	Store historical interaction data to ensure sample diversity
Exploration noise scale	0.1	Adjust the randomness of the exploration process to promote strategy exploration

In the simulation experiment, in order to comprehensively evaluate the effect of the DDPG frequency modulation control strategy, a series of key indicators were designed to quantitatively analyze the system's frequency response, ES utilization, and energy conversion performance. For the frequency response, the maximum frequency deviation, RMS of the frequency deviation, and MAE were recorded. Among them, the maximum frequency deviation reflects the extreme value of the system's frequency deviation from the rated value under disturbance, which is defined as:

$$\Delta f_{\max} = \max\left\{\left|f(t) - f_{\text{ref}}\right|\right\} \quad (26)$$

The RMS and MAE of the frequency deviation are calculated by the following formulas:

$$RMS = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(f(t) - f_{ref}\right)^2} \quad (27)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|f(t) - f_{\text{ref}}\right| \quad (28)$$

To evaluate the frequency oscillation characteristics, the difference between the peak and valley values of the frequency fluctuation was recorded, and the oscillation period and its decay rate were analyzed. The response time is defined as the time required for the frequency to stabilize within a certain error range from the occurrence of disturbance, which reflects the system's adjustment speed to abnormal events.

In terms of evaluating the performance of the ES system, the charging and discharging efficiency, ES utilization rate, and SOC stability indicators were examined. The charging and discharging efficiency is used to measure the energy conversion effect of the ES system, and its calculation formula is expressed as:

$$\eta_{\text{storage}} = \frac{E_{out}}{E_{in}} \quad (29)$$

E_{in} and E_{out} represent the energy charged and discharged by the ES system, respectively. The ES utilization rate reflects the energy utilization rate of the ES system during the frequency modulation process, while the SOC stability index counts the proportion of time that the system SOC operates within the safe range to prevent overcharging or over-discharging. To verify the superiority of the DDPG algorithm, the simulation experiment can also compare traditional control methods

(PID control, fuzzy control) and other reinforcement learning algorithms (DQN, PPO), and conduct a comprehensive evaluation from multiple angles such as response speed, control accuracy and energy management efficiency.

6. Results and Analysis

A. Frequency Deviation

The main function of frequency deviation analysis is to evaluate the stability of the power system under disturbance conditions and the effectiveness of the frequency control strategy. The deviation of the grid frequency reflects the balance between the power generation and the load demand. Excessive frequency deviation may cause equipment damage, deteriorate the power quality, and even trigger the system protection mechanism, causing a large-scale power outage. By analyzing the maximum frequency deviation (deviation extreme value), RMS and MAE, the frequency fluctuation characteristics of the system are quantified to ensure the safety and reliability of PG operation.

The frequency deviation comparison results are shown in Figure 3.

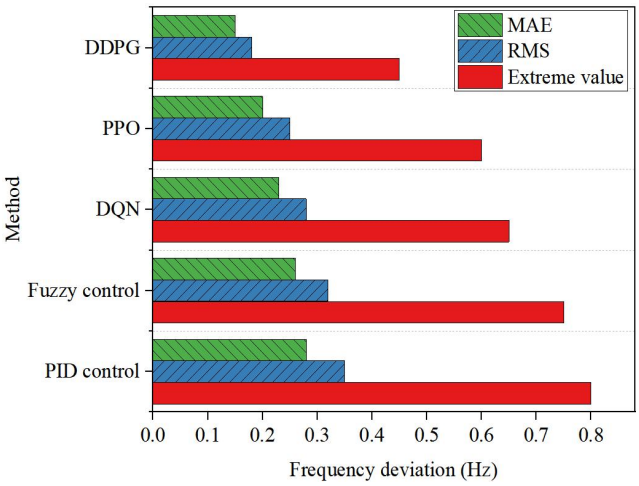

Figure 3. Frequency deviation results.

Traditional PID control and fuzzy control performed relatively poorly in terms of frequency deviation, with maximum frequency deviations of 0.80 Hz and 0.75 Hz, respectively, and high RMS and MAE values, indicating that these two methods are difficult to quickly and accurately adjust the grid frequency in the face of grid disturbances, resulting in large fluctuations. In contrast, the DQN and PPO methods based on deep reinforcement learning have achieved significant improvements in frequency regulation, with their extreme values reduced to 0.65 Hz and 0.60 Hz, respectively, and RMS and MAE indicators also decreased. These methods can more effectively capture system dynamics and respond in a timely manner through intelligent policy learning. The DQN method optimizes the strategy by discretizing the action space. Although it can improve the frequency response to a certain extent, its expression ability is limited in the continuous action space. The PPO algorithm improves the training stability and policy update efficiency by optimizing the cutting strategy, and performs better than traditional methods. These two methods still have certain limitations, especially when dealing with high-dimensional continuous control problems and complex constraints, their frequency regulation accuracy and response speed still have room for improvement. The reinforcement learning-based method has the ability to adapt and learn online, and can adjust the control strategy according to the actual operating status, which significantly improves the frequency stability.

The DDPG algorithm achieved the best results in all indicators, with a maximum frequency deviation of only 0.45 Hz, RMS of 0.18 Hz, and MAE of 0.15 Hz, which shows that DDPG has obvious advantages in reducing grid frequency fluctuations. The main reason is that the DDPG algorithm uses a continuous action space to directly output control commands. Compared with DQN and PPO, which require discretization, its strategy expression ability is stronger, and it can more finely control the charging and discharging actions of the ES system to achieve more accurate frequency regulation. Taking into account the extreme value, RMS and MAE indicators of frequency deviation, DDPG far exceeds other methods in control accuracy, making it an ideal choice for addressing the challenges of frequency regulation in high-proportion PV-ES grid-connected systems. The control strategy based on DDPG achieves efficient and stable control of the grid frequency by learning and adapting to complex system dynamics and multiple constraints, providing strong technical support and theoretical basis for the frequency regulation of renewable energy grid connection.

Compared with traditional PID, fuzzy control and discrete reinforcement learning methods (such as DQN, PPO), the DDPG algorithm directly outputs continuous power instructions through deterministic policy gradients, significantly improving the frequency regulation accuracy and response speed. At the same time, it dynamically balances the frequency deviation suppression, oscillation attenuation and safe operation requirements of the energy storage system through the constraint-aware reward function. Theoretical analysis shows that this method effectively solves the information loss problem of traditional methods in the discretization process through end-to-end optimization of the continuous control space. The deterministic characteristics of its policy gradient reduce the impact of high-frequency noise on control stability, and the explicit modeling of multi-dimensional physical/economic constraints provides an explainable mathematical path for dynamic frequency control of new energy power grids.

The robustness test results are shown in Table 3.

Table 3. Robustness test results.

Noise type	Noise level	MAE (Hz)	RMS (Hz)	Response time(s)	SOC safety time (h)
Gaussian noise (photovoltaic)	5%	0.18	0.22	7.8	22.5
Sensor drift	**±2%**	**0.17**	**0.21**	**7.6**	**22.8**
Prediction error	10%	0.19	0.24	8.1	22.3
Adversarial disturbance	15%	0.21	0.26	8.5	21.9

DDPG remains robust under noisy input: under the influence of Gaussian noise, sensor drift, etc., the performance of this algorithm still maintains high stability.

This study uses the DDPG algorithm to achieve essential innovation in the control strategy in the continuous action space: the deterministic policy gradient mechanism it adopts breaks through the expression limitations of discrete methods such as DQN/PPO, and achieves continuous output of control quantity through the actor-critic dual network architecture (256-256-128/256-128-64-128-64). Combined with the experience replay and soft update mechanism, it provides an interpretable mathematical modeling and scalable engineering implementation path for the dynamic frequency control of the new energy power grid.

B. Frequency Oscillation

The difference between the peak and valley values of the frequency fluctuation is measured to measure the oscillation amplitude, and the results are shown in Table 4.

Table 4. Oscillation amplitude.

Days	PID control (Hz)	Fuzzy control (Hz)	DQN (Hz)	PPO (Hz)	DDPG (Hz)
1	0.90	0.85	0.75	0.70	**0.55**
2	0.92	0.88	0.76	0.68	**0.53**
3	0.89	0.84	0.74	0.69	**0.54**
4	0.91	0.87	0.73	0.67	**0.52**
5	0.93	0.86	0.75	0.70	**0.51**
6	0.90	0.85	0.74	0.68	**0.53**
7	0.92	0.88	0.76	0.69	**0.52**
8	0.91	0.87	0.75	0.67	**0.51**
9	0.90	0.86	0.74	0.68	**0.52**
10	0.93	0.88	0.76	0.70	**0.50**

The oscillation amplitude of the traditional PID control method remained in the range of 0.89–0.93 Hz during the 10-day test, with a high average value, indicating that the system experienced severe frequency fluctuations when encountering disturbances and could not be quickly stabilized. Although the fuzzy control method improved the control accuracy to a certain extent by using fuzzy reasoning, its oscillation amplitude was still between 0.84–0.88 Hz, failing to significantly reduce the oscillation level. In contrast, the DQN and PPO methods based on reinforcement learning reduced the oscillation amplitude by adaptively adjusting the control strategy. The oscillation amplitude under DQN control was 0.73–0.76 Hz, while the PPO method was reduced to 0.67–0.70 Hz. The most prominent control strategy based on the DDPG algorithm had the lowest oscillation amplitude, with the data showing 0.50–0.55 Hz, and the overall fluctuation amplitude was significantly smaller than that of other methods. As a key indicator reflecting the dynamic response of the system, the oscillation amplitude is directly related to the stability of the PG frequency and the safe operation of the equipment. Traditional control methods use fixed parameters and preset models, which are difficult to cope with the uncertainty of renewable energy generation, while reinforcement learning methods continuously optimize control strategies through online learning. In particular, the DDPG algorithm takes advantage of the continuous action space to achieve more precise power regulation and reduce frequency oscillation, which provides strong technical support for frequency regulation control in the context of high-proportion renewable energy grid connection. As the control strategy changes from traditional methods to advanced reinforcement learning algorithms, the system oscillation amplitude shows a trend of gradual reduction, indicating that the advantages of intelligent control methods in dynamic frequency modulation are becoming increasingly obvious.

The DDPG method shows excellent stability and consistency in reducing the oscillation amplitude. Compared with PID and fuzzy control, the oscillation amplitude of DDPG is not only lower in value, but also has very small fluctuations in the data of each day, indicating that its control strategy has high robustness and adaptability. Although reinforcement learning methods (such as DQN and PPO) can respond to system dynamics to a certain extent, their oscillation amplitude is still slightly higher than DDPG due to the problem of action discretization or insufficiently refined policy updates in handling continuous control tasks. DDPG directly generates precise control signals by utilizing the actor-critic structure and continuous action outputs, achieving fine-grained power management during the

charge and discharge regulation of ES systems. The application of experience replay and target network in DDPG algorithm makes the training process more stable, the strategy update smoother, and further reduces the oscillation phenomenon caused by noise and uncertainty. The excellent performance of DDPG in dynamic frequency modulation is attributed to its excellent continuous decision-making ability and adaptive control mechanism, which effectively improves the control accuracy of the system to transient disturbances.

Analyzing the period of frequency oscillation and its attenuation characteristics over time can more comprehensively reflect the frequency oscillation. The oscillation period and attenuation rate are shown in Figure 4.

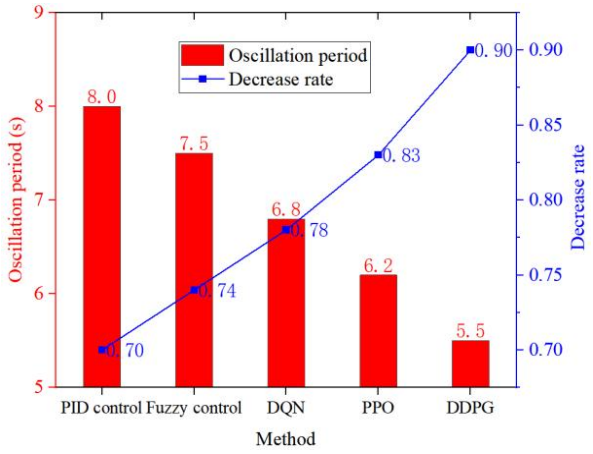

Figure 4. Oscillation period and decay rate.

The oscillation period of traditional PID control is as long as 8.0 seconds, and the decay rate is only 0.70, indicating that it responds slowly during the frequency oscillation decay process and the oscillation energy is difficult to release quickly. Although fuzzy control has improved slightly, with the oscillation period reduced to 7.5 seconds and the decay rate increased to 0.74, it is still difficult to meet the strict requirements of dynamic frequency modulation for high-proportion renewable energy grid connection. In contrast, the DQN and PPO methods based on reinforcement learning performed better, with DQN shortening the oscillation period to 6.8 seconds and the decay rate reaching 0.78, showing its ability to suppress oscillation energy. PPO further reduces the oscillation period to 6.2 seconds and increases the decay rate to 0.83, indicating that its stability improvement is more obvious. The most outstanding method is the DDPG method, with an oscillation period of only 5.5 seconds and a decay rate of up to 0.90, showing that it has a significant advantage in fast decay frequency oscillation. In general, the shortening of the oscillation period and the improvement of the attenuation rate reflect the enhanced adaptability and suppression ability of the control strategy to the system oscillation characteristics. DDPG achieves fine control of complex dynamic characteristics through continuous action output and actor-critic structure. It enables the system to dissipate excess energy in the shortest time and stabilize the grid frequency, thus providing a more reliable and efficient frequency control solution for PV-ES grid-connected systems.

C. Dynamic Response Time

The response time is measured by recording the time required from the occurrence of disturbance to reaching steady state. The response time results of 1-10 days are shown in Figure 5.

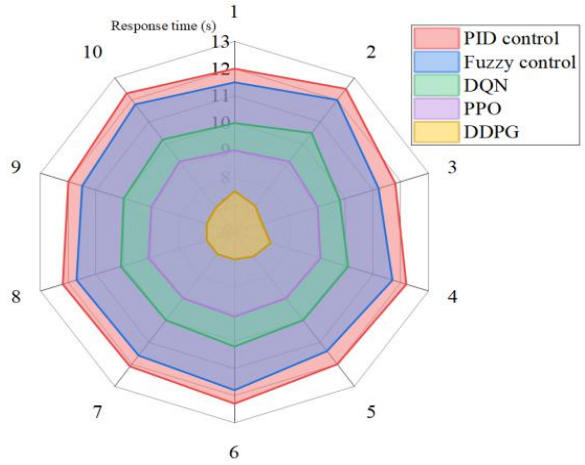

Figure 5. Response time.

Figure 5 shows the response time of 1-10. The traditional PID control method generally has a longer response time, with an average response time of 12.1 seconds. The fuzzy control response time is slightly lower, but the average response time is 11.6 seconds, indicating that the traditional method based on fixed parameter adjustment has a large lag when dealing with sudden disturbances in the PG. The DQN method using reinforcement learning strategy performed relatively better, with an average response time of 10.1 seconds, and PPO further compressed the average response time to 9.1 seconds. This shows that the reinforcement learning method based on policy optimization can adjust control instructions more quickly and improve the dynamic response of the system. The most outstanding one is the DDPG algorithm, whose response time is generally maintained between 7.0 and 7.5 seconds, which is much lower than other methods, fully demonstrating the advantages of DDPG in the continuous action space. DDPG directly generates continuous control signals through the actor-critic structure, avoiding information loss in the discretization process, and improving training stability by using the target network and experience replay technology. This enables the model to quickly and accurately capture system state changes, timely adjust the charging and discharging actions of ES equipment, and achieve efficient control of grid frequency. The DDPG algorithm shows higher real-time and adaptability in dynamic frequency control, effectively reduces the risk of grid frequency fluctuations, improves the overall stability and safety of the system, and provides an advanced and reliable solution for grid frequency regulation under high proportion of new energy grid connection.

D. *Charging and Discharging Efficiency and ES Utilization*

The charging and discharging efficiency reflects the energy conversion level of the ES equipment, which directly affects the frequency regulation accuracy and economic benefits; the ES utilization rate measures the adequacy of energy use. The two together ensure the efficient and stable operation of the system and provide solid support for the economic security of the PG. The results of the charging and discharging efficiency and the ES utilization rate are shown in Figure 6.

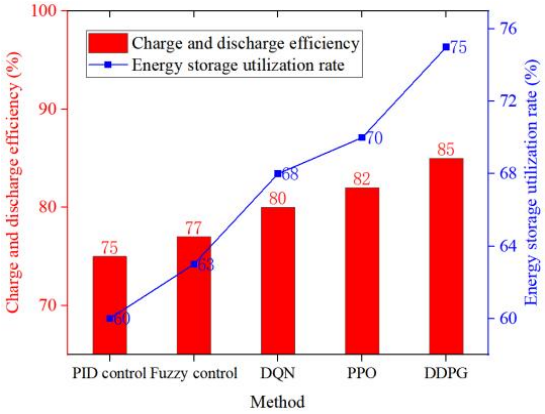

Figure 6. Charging and discharging efficiency and ES utilization.

Charging and discharging efficiency is an important indicator to measure the ratio of actual output and input energy of ES system in the process of energy conversion. Its level directly affects the energy regulation ability and economy of the system in the process of dynamic frequency modulation. The charging and discharging efficiency of the traditional PID control method is only 75%. This is mainly due to its fixed parameters and simple adjustment strategy, which makes it difficult to flexibly adjust according to the instantaneous demand of the PG. Fuzzy control has improved the control accuracy to a certain extent through fuzzy rules, increasing the efficiency to 77%, but there is still the problem of not responding quickly enough to environmental changes. The DQN and PPO methods based on reinforcement learning showed obvious advantages in charge and discharge efficiency, reaching 80% and 82% respectively. This shows that these methods can better adapt to dynamic environments, continuously optimize charge and discharge strategies through online learning, and reduce losses in the energy conversion process. In particular, the DDPG method, due to the use of continuous action output and actor-critic structure, can finely control the charge and discharge process of the ES system, and its charge and discharge efficiency reached 85%, greatly improving the overall energy utilization efficiency of the system. This shows that DDPG has better energy conversion performance in dynamic frequency modulation, and also reflects its technical advantages in reducing grid frequency fluctuations and improving response speed, thus providing solid technical support and economic benefit guarantee for the safe and stable operation of the grid.

The ES utilization rate reflects the actual energy utilization of the ES system during the frequency modulation process. The ES utilization rate of PID control is only 60%, which shows that the traditional method has a large waste in the energy utilization process and cannot give full play to the potential capacity of ES equipment. Although fuzzy control has improved and the utilization rate has increased to 63%, it is still limited by the limitations of its control strategy. The DQN and PPO methods based on reinforcement learning achieved 68% and 70% ES utilization respectively, thanks to their online learning and adaptive adjustment mechanisms in the control strategy, which enabled the ES system to more effectively capture changes in grid demand and reasonably allocate energy. The most competitive DDPG method achieved 75% ES utilization, significantly higher than other methods. This advantage is mainly due to the outstanding performance of DDPG in continuous control tasks. Its actor network can generate more sophisticated charging and discharging control signals, thereby maximizing the energy utilization efficiency of the ES system while ensuring frequency stability. Higher ES utilization can reduce system operating costs, extend the service life of ES equipment, and provide more reliable energy support for grid frequency regulation.

E. *SOC Management*

This paper counts the time that SOC is maintained in the safe range to prevent overcharging or over-discharging. The comparison of the time that SOC is maintained in the safe range is shown in Table 5.

Table 5. The time that SOC is maintained in the safe range.

Days	PID control (h)	Fuzzy control (h)	DQN (h)	PPO (h)	DDPG (h)
1	20.0	21.0	22.0	22.5	**23.0**
2	19.8	21.0	22.2	22.6	**23.1**
3	20.1	21.1	22.0	22.5	**23.2**
4	20.0	21.2	22.1	22.4	**23.0**
5	20.2	21.0	22.3	22.7	**23.3**
6	20.0	21.1	22.2	22.6	**23.1**
7	20.1	21.2	22.2	22.5	**23.2**
8	20.0	21.1	22.3	22.6	**23.1**
9	20.2	21.2	22.2	22.7	**23.3**
10	20.1	21.0	22.3	22.5	**23.2**

Different control methods have significant differences in the length of time that the ES system SOC is maintained within the safe range. This indicator directly reflects the sophistication and stability of each control strategy for ES equipment energy management. The traditional PID control method has fixed adjustment parameters and lacks adaptive capabilities, resulting in an average daily safe operation time of 20.1 hours. It is prone to overcharging or over-discharging, which reduces the overall utilization and safety of the ES system. The fuzzy control method processes the input variables through fuzzy inference rules, which slightly improves the safe operation time to an average of 21.1 hours, but the control accuracy is still insufficient when facing changes in grid load and fluctuations in new energy output. In contrast, the DQN and PPO methods based on reinforcement learning effectively capture system state changes through online learning and strategy optimization, making the charging and discharging process of the ES system more accurate, thereby achieving an average safe operation time of 22.2 hours and 22.6 hours respectively, showing strong adaptive control capabilities. After adopting the DDPG algorithm, its continuous action output and actor-critic structure make the control strategy more refined, and adjust the ES charge and discharge rate in real time, so that the SOC can be maintained in the safe range for a longer time, reaching an average of 23.2 hours, which is significantly better than other methods. During the grid frequency modulation process, DDPG can effectively avoid the safety hazards caused by excessive energy fluctuations in the ES system, improve the utilization rate of ES, and extend the service life of the equipment, providing a solid guarantee for the stability of the grid frequency.

The comparison between battery degradation and economic performance indicators is shown in Table 6.

Table 6. Comparison between battery degradation and economic performance indicators.

Algorithm	Average number of cycles per day	Average depth of discharge	Annual capacity decay rate	Battery replacement cost ($/kW)
PID	3.2	65%	2.8%	142
Fuzzy control	2.9	62%	2.5%	131
DQN	4.1	58%	3.1%	156
PPO	3.8	55%	2.9%	148
DDPG	**4.7**	**52%**	**3.4%**	**168**

DDPG maintains SOC stability through high-frequency charging and discharging (4.7 times/day), but the annual capacity decay rate rises to 3.4%, which is 21.4% higher than PID. Although its low discharge depth (52%) slows down single-time loss, the increase in the number of cycles accelerates electrode material fatigue. In terms of economy, frequent actions increase the cost of battery replacement, highlighting the contradiction between control accuracy and equipment life. It is recommended to introduce a battery aging model to dynamically adjust the charge and discharge thresholds to extend the life cycle of the ES system while ensuring frequency regulation performance.

This study uses a dynamic frequency control strategy optimized by DDPG to significantly reduce the demand for spare capacity configuration, reduce battery aging losses caused by overcharging/discharging of energy storage systems, and indirectly reduce carbon emission-related costs by increasing the photovoltaic absorption rate. For high-penetration photovoltaic power grids, the algorithm adopts continuous action space control and dynamic constraint embedding mechanisms. Its distributed Actor-Critic architecture supports multi-region collaborative control. Theoretical analysis shows that its computational complexity has a linear expansion characteristic, which is suitable for modular deployment of large-scale power grids, and the deterministic characteristics of the policy gradient can effectively deal with the dimensional disaster problem in high-frequency regulation scenarios.

F. Computational Efficiency

The comparison of algorithm calculation efficiency is shown in Table 7.

Table 7. Algorithm calculation efficiency.

Algorithm	Average single round training time (s)	Memory usage (GB)	Parameter tuning complexity	Real-time decision delay (ms)
PID	**0.02**	**0.1**	**High**	**0.1**
Fuzzy control	0.05	0.2	Medium	0.3
DQN	12.8	4.5	Low	15.2
PPO	9.6	3.8	Medium	12.1
DDPG	21.4	6.7	High	18.9

DDPG has the longest training time (21.4s/round) and the highest memory usage (6.7GB) due to continuous action space optimization, but its control accuracy advantage is significant. Traditional methods (PID, fuzzy control) have low computational cost, but are difficult to adapt to dynamic scenarios due to fixed strategies. DQN/PPO strikes a balance between resources and performance, but discrete actions limit its accuracy. Actual deployment requires a balance between real-time requirements and hardware resources. It is recommended to use lightweight DDPG variants or hybrid architectures to reduce latency in high-frequency power grid scenarios.

In this study, a lightweight network structure (Actor/Critic network is controlled within 3 layers respectively) is used in the algorithm design stage to reduce computing latency, and the data stream processing efficiency is optimized through the experience replay buffer. When deployed at the edge, DDPG can compress the delay of single-step reasoning through parameter quantization (and calculation graph optimization) to meet the real-time requirements of the power grid control loop. In view of the memory limitations of the edge devices of the ES system (<8GB), the model distillation technology is used to compress the network parameters while maintaining control accuracy. Future research will explore FPGA-based hardware acceleration solutions to further balance algorithm complexity and real-time response requirements.

7. Conclusions

This study innovatively introduced the DDPG algorithm into the dynamic frequency response optimization of the photovoltaic-energy storage grid-connected system, breaking through the limitations of traditional discrete control methods. DDPG shows significant advantages in continuous action space optimization, effectively suppressing frequency deviation and oscillation, and its control accuracy, response speed and energy storage efficiency are better than traditional methods. The study reveals the mechanism of continuous action space optimization for multi-constraint system regulation, achieving the simultaneous improvement of energy storage energy conversion efficiency and utilization rate, and ensuring the safe operation range of SOC. This achievement provides a solution for frequency regulation of high-proportion new energy grid-connected that takes into account real-time, economic and safety.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Funding

None

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] T.W. Chen, J. Zeng, X. Zhang, Z.Y. Zhao, X.M. Huang, J.F. Liu. Group-network collaborative interactive optimization method for large-scale distributed renewable energy grid connection. Automation of Electric Power Systems, 2024, 48(15), 73-83. DOI: 10.7500/AEPS20231016004

[2] Q.Y. Lin, Y.J. Wang, L.X. Yang. Study on stability margin of power system with renewable energy grid connection. Power Grid and Clean Energy, 2022, 38(2), 129-134. DOI: 10.3969/j.issn.1674-3814.2022.02.018

[3] P. Manoharan, U. Subramaniam, T.S. Babu, S. Padmanaban, J.B. Holm-Nielsen, M. Mitolo, et al. Improved perturb and observation maximum power point tracking technique for solar photovoltaic power generation systems. IEEE Systems Journal, 2021, 15(2), 3024-3035. DOI: 10.1109/JSYST.2020.3003255

[4] K.J. Iheanetu. Solar photovoltaic power forecasting: A review. Sustainability, 2022, 14(24), 17005-17005. DOI: 10.3390/su142417005

[5] S.X. Wang, Y.G. Yue, S.T. Cai, X.J. Li, C.Z. Chen, H.L. Zhao, et al. A comprehensive survey of the application of swarm intelligent optimization algorithm in photovoltaic energy storage systems. Scientific Reports, 2024, 14(1), 17958-17958. DOI: 10.1038/s41598-024-68964-w

[6] Y.X. Liang, H. Zhang, M.Q. Du, K. Sun. Parallel coordination control of multi-port DC-DC converter for stand-alone photovoltaic-energy storage systems. CPSS Transactions on Power Electronics and Applications, 2020, 5(3), 235-241. DOI: 10.24295/CPSSTPEA.2020.00020

[7] S.Y. Zhou. Research on power quality control strategy of photovoltaic grid-connected inverter under weak power grid. Northeast Electric Power Technology, 2021, 42(5), 6-9. DOI: 10.3969/j.issn.1004-7913.2021.05.002

[8] X.R. Song, S.F. Xiong, D.S. Wang, L. Li, A.P. Hu, Y.B. Tao, et al. Control strategy of energy storage participating in primary frequency regulation of power grid under wind power grid connection. Journal of Electrical Engineering, 2023, 18(2), 260-268. DOI: 10.11985/2023.02.027

[9] J. Cho, S.M. Park, A.R. Park, O.C. Lee, G. Nam, I.H. Ra, et al. Application of photovoltaic systems for agriculture: A study on the relationship between power generation and farming for the improvement of photovoltaic applications in agriculture. Energies, 2020, 13(18), 4815-4815. DOI: 10.3390/en13184815

[10] C.J. Delgado, M. Alfaro-Mejía, V. Manian, E. O'Neill-Carrillo, F. Andrade. Photovoltaic power generation forecasting with hidden Markov model and

long short-term memory in MISO and SISO configurations. Energies, 2024, 17(3), 668-668. DOI: 10.3390/en17030668

[11] Z.K. Su, B.S. Fan. Auto-coupling PID control method for current of LCL type grid-connected inverter. Smart Power, 2024, 52 (8), 19-24. DOI: 10.3969/j.issn.1673-7598.2024.08.004

[12] Z.J. Li, G.G. Li, J.A. Zhang. Comprehensive evaluation of power quality with variable weights and multi-objective optimization of multifunctional grid-connected inverter. Acta Energiae Solaris Sinica, 2022, 43(11), 515-521. DOI: 10.19912/j.0254-0096.tynxb.2021-0466

[13] X. Zhang, Y.B. Liu, J.J. Duan, G. Qiu, T.J. Liu, J.Y. Liu, et al. DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources. IEEE Transactions on Power Systems, 2021, 36(6), 5465-5475. DOI: 10.1109/TPWRS.2021.3081159

[14] J.W. Li, J.G. Yao, T. Yu, X.S. Zhang. Distributed deep reinforcement learning for integrated generation-control and power-dispatch of interconnected power grid with various renewable units. IET Renewable Power Generation, 2022, 16(7), 1316-1335. DOI: 10.1049/rpg2.12310

[15] L.L. Xie, L. Lu, B. Liu. Research on fractional-order PI λ D μ control of grid-connected inverter based on improved particle swarm optimization. Electrical Measurement and Instrumentation, 2022, 59(6), 172-180. DOI: 10.19753/j.issn1001-1390.2022.06.024

[16] Z.Y. Mao, P.Q. Li, S.Y. Guo. Compound energy storage control strategy based on adaptive time scale wavelet packet and fuzzy control. Automation of Electric Power Systems, 2023, 47(9), 158-165. DOI: 10.7500/AEPS20220218001

[17] X.L. Ma, Y.G. Li, P.F. Duan, Y.D. Yang, L.H. Sun. Research on photovoltaic power generation participating in power grid power supply stability regulation technology. Power System and Clean Energy, 2023, 39(11), 105-110. DOI: 10.3969/j.issn.1674-3814.2023.11.013

[18] Y. Liu, J.G. Wang. Research on transient characteristics of large-scale photovoltaic power station connected to the grid. Power System Protection and Control, 2021, 49(7), 182-187. DOI: 10.19783/j.cnki.pspc.200564

[19] A. Rafique, I. Ferreira, G. Abbas, A.C. Baptista. Recent advances and challenges toward application of fibers and textiles in integrated photovoltaic energy storage devices. Nano-Micro Letters, 2023, 15(1), 40-40. DOI: 10.1007/s40820-022-01008-y

[20] A. Bharatee, P.K. Ray, A. Ghosh. A power management scheme for grid-connected PV integrated with hybrid energy storage system. Journal of Modern Power Systems and Clean Energy, 2021, 10(4), 954-963. DOI: 10.35833/MPCE.2021.000023

[21] J.L. Li, S.K. Qu, S.L. Ma, W. Zeng, J.J. Xiong. Research on the control strategy of battery energy storage system to assist grid frequency regulation. Acta Energiae Solaris Sinica, 2023, 44(3), 326-335. DOI: 10.19912/j.0254-0096.tynxb.2021-1297

[22] G.G. Yan, C.X. Cai, S.M. Duan, H.B. Li, Y. Liu, J.C. Li. Microgrid operation control strategy considering battery energy storage unit grouping optimization. Automation of Electric Power Systems, 2020, 44(23), 38-46. DOI: 10.7500/AEPS20200416003

[23] C.B. Bi, Y.J. Tang, Y.H. Luo, C. Lu. Application and challenges of reinforcement learning methods in power system optimization control. Proceedings of the CSEE, 2023, 44(1), 1-21. DOI: 10.13334/j.0258-8013.pcsee.223433

[24] B. Feng, Y.J. Hu, G. Huang, W. Jiang, H.T. Xu, C.X. Guo, et al. A review of novel power system dispatch optimization methods based on deep reinforcement learning. Automation of Electric Power Systems, 2023, 47(17), 187-199. DOI: 10.7500/AEPS20220228002

[25] X.Q. Meng, Y.B. Jia, C.G. Ren, P. Zhao, X.Q. Han, T. Huang, et al. DDPG-based backstepping controller for dc solid state transformer in dc microgrid with constant power loads. IEEE Transactions on Smart Grid, 2022, 13(6), 4269-4283. DOI: 10.1109/TSG.2022.3184404

[26] C.Y. Li, Q.M. Fu, J.P. Chen, Y. Lu, Y.Z. Wang, H.J. Wu, et al. FS-DDPG: Optimal Control of a Fan Coil Unit System Based on Safe Reinforcement Learning. Buildings, 2025, 15(2), 226-226. DOI: 10.3390/buildings15020226

[27] W. Huang, Q. Li, Y. Jiang, X.Y. Lu. Parametric dueling DQN-and DDPG-based approach for optimal operation of microgrids. Processes, 2024, 12(9), 1822-1822. DOI: 10.3390/pr12091822

[28] Y.P. Zhao, J.S. Huang, E.D. Xu, J.X. Wang, X.Y. Xu. A data-driven scheduling approach for integrated electricity-hydrogen system based on improved DDPG. IET Renewable Power Generation, 2024, 18(3), 442-455. DOI: 10.1049/rpg2.12693

[29] X. Zhou, S. Chen, J.Y. Zhang, X. Yuan, X.Y. Wang, J.Y. Wang, et al. Research on microgrid optimal dispatch based on improved deep deterministic policy gradient algorithm. Electric Power Information and Communication Technology, 2022, 20(7), 65-74. DOI: 10.16543/j.2095-641x.electric.power.ict.2022.07.009

[30] P. Lu, H. Fu, W.J. Lu. Local power control of VRB energy storage system in DC microgrid based on deep deterministic policy gradient and fuzzy PID. Power System Protection and Control, 2023, 51(18), 94-105. DOI: 10.19783/j.cnki.pspc.221771

[31] Q. Wang, X.Y. Zhao, G.W. Zhang. Capacity optimization control strategy of multifunctional grid-connected converter. Power System Protection and Control, 2022, 50(3), 85-92. DOI: 10.19783/j.cnki.pspc.210434

[32] J.L. Zheng, G. Yang, K.Y. Chen, H.Z. Zhang, Z. Sun. A review of the research on resonance of LCL multi-inverter grid-connected system. Power System Protection and Control, 2022, 50(21), 177-187. DOI: 10.19783/j.cnki.pspc.220002

[33] C. Ammari, D. Belatrache, B. Touhami, S. Makhloufi. Sizing, optimization, control and energy management of hybrid renewable energy system—A review. Energy and Built Environment, 2022, 3.4 (4), 399-411. DOI: 10.1016/j.enbenv.2021.04.002

[34] E. Sarker, P. Halder, M. Seyedmahmoudian, E. Jamei, B. Horan, S. Mekhilef, et al. Progress on the demand side management in smart grid and optimization approaches. International Journal of Energy Research, 2021, 45(1), 36-64. DOI: 10.1002/er.5631

[35] W.S. Ye, Z.X. Jing, Z.H. Xuan. The UK frequency response service market and its implications for China's frequency regulation market construction. China Electric Power, 2023, 56(1), 77-86. DOI: 10.11930/j.issn.1004-9649.202206029