EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Voltage Fluctuation Control Strategy Based on Reinforcement Learning Digital Twin Model for Wind-Solar-Water-Storage Smart Microgrid Energy Storage Converter Side

Erbao Yan*

Guangzhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Guangzhou, 510800, China

Abstract

With the increasing penetration of distributed renewable energy (DRE), the stable operation of smart microgrids integrating wind, solar, hydro, and storage faces significant challenges. The core issue lies in the strong intermittency and random-ness of distributed generation outputs, which can easily trigger voltage fluctuations, particularly at the grid connection points of Power Conversion Systems (PCS), seriously threatening power quality and supply reliability. This study innovatively integrates Deep Reinforcement Learning and Digital Twin (DRL+DT) technologies to establish a novel control framework, which constructs a digital twin of the microgrid. Based on the Proximal Policy Optimization (PPO) algorithm, a DRL controller is designed, with clearly defined state spaces, action spaces, and a composite reward function incorporating multi-objective constraints. This modeling approach transforms the voltage control problem into a sequential decision-making task solvable through data-driven methods. The trained policy neural network serves as an intelligent controller for performance comparison in subsequent sections. To validate the proposed approach, simulation tests verify that the proposed method suppresses voltage fluctuations more rapidly and smoothly compared to conventional PI control and single DRL approaches. The voltage deviation is reduced by 62.4%, while the State of Charge (SOC) of the energy storage system is effectively maintained within a healthy range. The experimental results not only verify the technical advantages of the combined DRL+DT framework in addressing complex energy control challenges in microgrids but also demonstrate its capability to support the development of highly resilient and self-healing smart microgrid control systems, highlighting both advanced functionality and practical utility.

Keywords: New energy, Microgrid, Voltage fluctuation control, Deep reinforcement learning, Proximal policy optimization

Received on 20 November 2025, accepted on 20 December 2025, published on 15 April 2026

doi: 10.4108/ew.12110

1. Introduction

The increasing penetration of distributed renewable energy (DRE) in microgrids has highlighted the critical importance of managing the high stochasticity and multi-timescale coupling effects on both the generation and load sides [1]. Conventional methods such as PI control, model predictive control (MPC), and model-free adaptive control (MFAC) exhibit limitations in control accuracy, and it remains challenging to achieve globally optimal coordinated control using a single approach. Consequently, there is a pressing need to enhance the applicability and adaptability of control models [2]. Current knowledge lacks a control paradigm capable of operating without reliance on precise mathematical models, underscoring the necessity for in-depth

* ervbi9ghs479966@163.com

research into adaptive learning mechanisms that can capture system dynamics in real time [3].

Recent studies in power management have explored various strategies. Smith, A. B. et al. proposed a droop-control-based voltage regulation method for microgrids, and incorporated active power, voltage and reactive power, and voltage droop characteristics [3]. A key limitation, however, is its strong dependence on accurately tuned droop coefficients and the inherent inability of static droop characteristics to adapt to dynamic system changes. In Reference [4], a model predictive control (MPC) strategy was designed for energy storage management to achieve anticipatory optimization; nevertheless, its performance is highly reliant on prediction model accuracy, and significant model mismatch can occur under highly fluctuating renewable power injection, leading to performance degradation. A robust control approach was introduced in Reference [5] to handle system uncertainties through an optimized Controller $^{\text{min-max}}$, albeit at the expense of performance under normal operating conditions to ensure robustness in worst-case scenarios. Schmidt, P. et al. applied an improved fuzzy logic control to converter regulation using an expert-knowledge-based fuzzy rule base [6]. However, the methodology for constructing rule sets and tuning membership functions lacks systematic rigor and offers limited optimization. The study in Reference [7] investigated a distributed cooperative control framework with an information exchange protocol among neighboring nodes, though communication delays and reliability issues adversely affect system stability. In Reference [8], a supervised learning-based neural network control was developed featuring an offline-trained feedforward controller. Its drawbacks include the requirement for extensive labeled data covering all operational scenarios and the lack of online adaptability to unknown dynamics. While these studies have contributed to improving system stability, they often involve trade-offs among model accuracy, computational complexity, and adaptive capability. None fully resolve the core challenge of achieving real-time, optimal, or autonomous voltage control under high uncertainty [9].

A methodology integrating deep reinforcement learning (DRL) with digital twin technology is proposed in this study. A high-fidelity digital twin model is constructed to emulate the dynamic behavior of the physical microgrid, which serves as a safe and realistic virtual environment for DRL agent training. A smart controller based on the proximal policy optimization (PPO) algorithm is designed, with carefully formulated state and action spaces and a composite reward function. This enables the DRL agent to autonomously learn an optimal control policy through trial and interaction within the digital twin. The strategy aims to dynamically coordinate the active and reactive power outputs of energy storage converters in real time. The proposed approach is anticipated to significantly enhance the dynamic performance of voltage regulation, not only enabling rapid fluctuation suppression but also ensuring the operational sustainability of the energy storage system (ESS). Ultimately, it is expected to achieve adaptive and robust control performance superior to conventional methods, without requiring an exact mathematical model.

The technical innovations of this study are listed here:

(1) An integrated framework of deep integration of DRL training and digital twin simulation is constructed to provide a feasible engineering path for key energy applications.

(2) A multi-objective collaborative optimization mechanism with a compound reward function is innovatively designed to guide agents to make optimal decisions that take into account both short-term effectiveness and long-term sustainability under multiple constraints.

(3) An intelligent coordinated control strategy is developed to achieve four-quadrant operation of power conversion systems (PCS) in energy storage systems (ESS), thereby unlocking their potential as flexible regulation resources. This approach significantly enhances the overall control flexibility and optimization scope of the system.

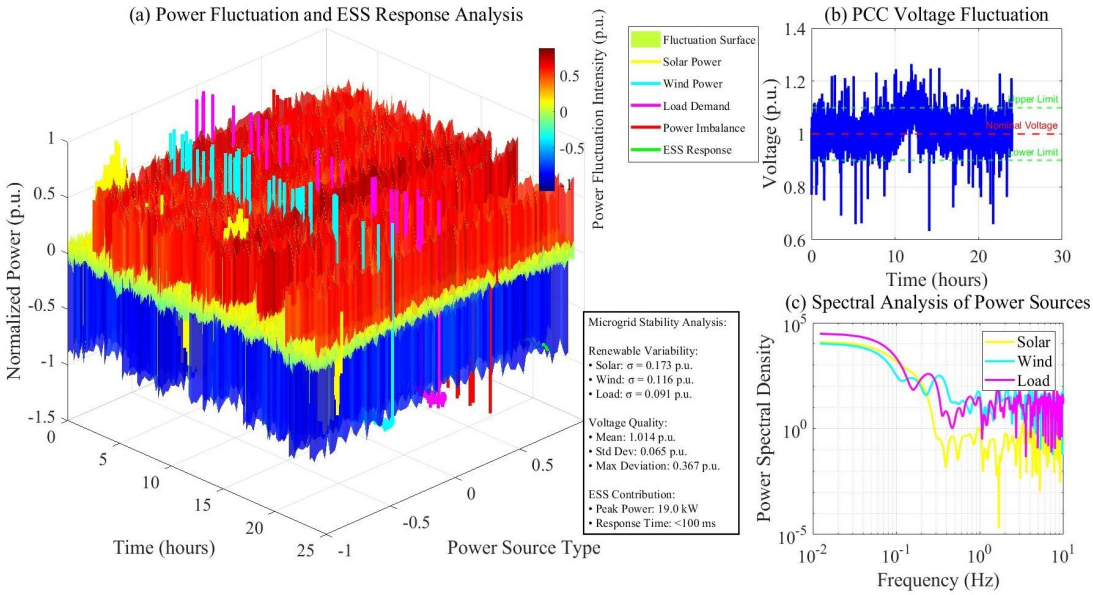

Fig. 1. The microgrid traditional control model

2. The relative work

2.1. Overall analysis of voltage fluctuation control in microgrid

As a critical platform for integrating distributed renewable energy (DRE), the smart microgrid faces core operational stability challenges stemming from the strong intermittency of primary energy sources such as wind and solar, as well as the stochastic nature of load demand [10]. The combined effect of these two uncertainty factors frequently disrupts real-time power balance within the system, thereby inducing sustained fluctuations and deviations in the voltage at the Point of Common Coupling (PCC) [11]. The Energy Storage System (ESS), which injects or absorbs active and reactive power via the Power Conversion System (PCS), is recognized as one of the most effective solutions for mitigating such power disturbances and supporting voltage stabilization [12]. A conventional control architecture of a microgrid is illustrated in Figure 1.

Therefore, the control strategy design of PCS has become the research focus to ensure the power quality and operation reliability of microgrid [13]. Current research as a whole faces the need to shift from traditional control relying on linear models to data-driven and intelligent decision-making paradigms dealing with highly nonlinear systems [14].

2.2. Comparative analysis of mainstream control technologies

In the field of microgrid voltage control, prevailing techniques exhibit distinct characteristics. PI control is simple and reliable, but lacks adaptive capability. Model Predictive Control (MPC) offers strong optimization performance yet is constrained by model accuracy [15]. Fuzzy logic control handles uncertainties effectively but relies heavily on expert knowledge for design, and Model-Free Adaptive Control (MFAC) eliminates the need for an explicit model but suffers from theoretical limitations in convergence and generalization [16]. The technical analysis is presented in Table 1.

Table 1. Comparison and analysis of mainstream technologies of microgrid voltage control.

Control Technology	Advantages	Disadvantages
PI Control [17]	Simple structure and the low calculation cost	Rely on the exact linear model, and the robustness is weak.
MPC [18]	A rolling optimization mechanism is adopted to deal with multi-variable constraints.	Heavy computational burden, heavily dependent on model accuracy
FLC [19]	It does not rely on accurate models and is good at dealing with nonlinear problems.	Design relies on expert experience and lacks self-learning ability.
MFAC [20]	Only I/O data is required.	Theoretical guarantees of convergence and stability require strict assumptions.
DRL [21]	Learning optimal strategies to handle highly nonlinear problems	Training needs a lot of data, and there is instability.

Because the inherent drawbacks of a single technology cannot be completely and perfectly solved, it is necessary to combine multiple methods for common optimization to meet the system requirements, as shown in Table 2.

Table 2. Comparison and analysis of technical integration scheme.

Combining technology	Way of implementation	Advantages	Limitations
PID-FLC combination [22]	Using FLC to tune PID parameters on line	Improve the adaptive ability of traditional PID.	Performance is limited by PID framework.
MPC-state estimator combination [23]	Kalman filter and other algorithms are used to provide state prediction for MPC.	Enhance the ability of MPC to cope with measurement noise and uncertainty.	Unresolved model dependency
DRL-simulation environment integration [24]	Training with high fidelity simulation model instead of real environment	Addressing safety risks in DRL physical training	Final performance depends on model and algorithm efficiency.

Combining technology	Way of implementation	Advantages	Limitations
Deep integration of DRL and DT	The Digital Twin provides a high-fidelity training environment and the DRL provides intelligent decision-making.	Both environmental authenticity and decision intelligence	High accuracy requirement for digital twin model

A technical pathway integrating deep reinforcement learning (DRL) and digital twin (DT) technology is adopted, in which a digital twin provides a high-fidelity virtual representation that closely approximates the physical entity. The dynamic characteristics of the system are described by a set of differential-algebraic equations. For a microgrid incorporating an Energy Storage System (ESS), the key dynamics on the AC side are used to construct the core electrical model of the digital twin.

3. Voltage control method of energy storage converter based on DRL+DT fusion

To address the issue of voltage fluctuation suppression in the Power Conversion System (PCS) within a renewable-integrated microgrid, comprising wind, photovoltaic, hydro, and storage sources, this study focuses on enhancing voltage quality at the Point of Common Coupling (PCC). The PCC voltage is subject to multiple disturbances, including fluctuations from wind and photovoltaic power generation, as well as random load variations [25].

3.1. Brief description of research object and technical framework

An innovative control architecture is proposed that deeply integrates DRL and DT technologies. This framework is designed to accurately simulate the dynamic responses of the physical system. The resulting control strategy is deployed to the central controller of the physical microgrid to enable real-time closed-loop control of the PCS. The overall system architecture is depicted in Figure 2.

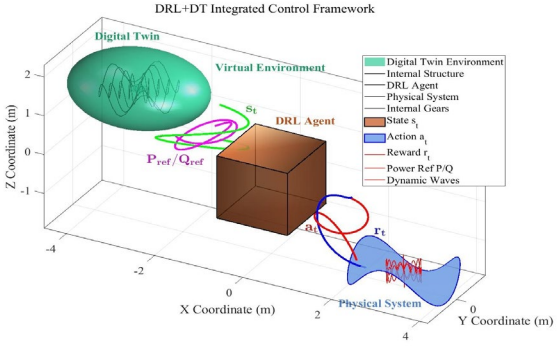

Fig. 2 The technical framework of the DRL+DT integrated control

Figure 1 shows the information interaction and closed-loop relationship between the digital twin environment, the DRL agent, and the physical system. The modeling accuracy of digital twins is the cornerstone of the effectiveness of the DRL strategy. Its construction needs to cover the dynamic characteristics of distributed generation, network topology, load, and ESS. Each component model constitutes the interactive environment E of the DRL agent.

3.2. Digital twin environment modeling

The randomness of wind-solar power and load is the core disturbance source. Its model usually adopts a time series modeling method. The output power model of the PV array can be expressed as:

$$P_{pv}(t) = P_{std} \cdot \frac{G(t)}{G_{std}} \cdot [1 + k \cdot (T_c(t) - T_{std})] \quad (1)$$

Where: $P_{pv}(t)$ is the output active power (W) of the PV array at time t; P_{std} is the rated power (W) under standard test conditions; $G(t)$ is the actual irradiance (W/m²) on the surface of the PV panel at time t; G_{std} is the irradiance (1000 W/m²) under standard test conditions; k is the power temperature coefficient (%/℃); $T_c(t)$ is the temperature (℃) of the PV cell at time t; T_{std} is the temperature at standard test conditions (25℃).

The linearized model is expressed as:

$$\Delta \mathbf{V} = \mathbf{J} \cdot \Delta \mathbf{P} + \mathbf{K} \cdot \Delta \mathbf{Q} \quad (2)$$

In the formula, $\Delta \mathbf{V}$ represents the N-dimensional column vector (V) formed by the voltage deviation value of each node; $\Delta \mathbf{P}$, $\Delta \mathbf{Q}$ represent the N-dimensional column vector (W, Var) formed by the deviation value of the active and reactive power injected by each node; $\mathbf{J}$, $\mathbf{K}$ represent the system active-voltage and reactive-voltage sensitivity matrix (V/W, V/Var), whose elements are determined by the network topology parameters and the initial operating point.

3.3. Design of deep reinforcement learning algorithm

In this paper, PPO (Proximal Policy Optimization) algorithm is used, and the interaction process is modeled as MDP (Markov Decision Process). It comprises a state space, an action space and a reward function.

The state space S shall comprehensively represent the system operation state, and the formula is:

$$s_t = [V_{pcc}, P_{load}, Q_{load}, P_{pv}, P_{wind}, SOC]^T \quad (3)$$

In the formula: s_t is the state vector at time t; V_{pcc} is the measured value of PCC point voltage (V); P_{load}, Q_{load} is the active and reactive power of the total load (W, Var); P_{pv}, P_{wind} is the active power of photovoltaic and wind turbine output (W); and SOC is the current state of charge of ESS (%).

The action space A is defined as the control instructions of the agent to the PCS. Its output is a continuous value:

$$a_t = [\Delta P_{ref}, \Delta Q_{ref}]^T \quad (4)$$

Where: a_t denotes the action vector at time t; $\Delta P_{ref}, \Delta Q_{ref}$ denotes the adjustment quantity (W, Var) of the active and reactive power reference values at time t.

The reward function R designs a multi-objective compound reward function to stabilize the voltage and ensure the healthy operation of the energy storage at the same time:

$$r_t = -\alpha \cdot (V_{pcc} - V_{ref})^2 - \beta \cdot (SOC - SOC_{opt})^2 - \gamma \cdot (\Delta P_{ref}^2 + \Delta Q_{ref}^2) \quad (5)$$

In it: r_t is the instant reward obtained at time t; V_{ref} is the rated reference value (V) of PCC voltage; SOC_{opt} is the optimal working state of charge of ESS (such as 60%); α, β, γ are the weight coefficients of each penalty factor, which are used to balance the importance of different optimization objectives.

The goal of the PPO algorithm is to find a strategy $\pi_\theta(a_t \mid s_t)$ that maximizes the expected cumulative discounted reward:

$$\eta(\pi_\theta) = \mathbb{E}\tau \sim \pi\theta[\sum_{t=0}^{T} \gamma^t r_t] \quad (6)$$

Where, τ is the trajectory from the initial state to the terminal state, γ represents the discount factor, and θ represents the strategy neural network parameter. Its core update formula guarantees stability by tailoring the surrogate target:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\text{ò}, 1+\text{ò})\hat{A}_t)] \quad (7)$$

In the formula: $L^{CLIP}(\theta)$ represents the pruning loss function under the parameter θ; $r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{old}}(a_t \mid s_t)}$ represents the probability ratio of the new and old strategies; $\hat{A}_t$ represents the estimated value of the advantage function at time t; ò represents the pruning hyperparameter, which is used to limit the amplitude of each strategy update.

Algorithm pseudo code:

Algorithm 1: PPO-based Control for Energy Storage PCS

Input: Environment E, hyperparameters $\{\alpha, \gamma, \varepsilon, \lambda\}$, max episodes E

Output: Optimized policy π_θ

1.Initialize policy network π_θ and value network V_ϕ

2.for episode = 1 to E do

3. Reset environment: $S_0 \leftarrow E.\text{init}(\cdot)$

4. for $t = 0$ to T do

5. Sample action: $\alpha \sim \pi_\theta(\cdot \mid S)$

6. Execute: $(s_{+1}, r_t, \text{done}) \leftarrow E.\text{step}(\alpha)$

7. Store transition: (s, α, r, s_{+1})

8. end for

9. Compute advantages: $\hat{A} \leftarrow \text{GAE}(\gamma, \lambda)$

10.Update policy: $\theta \leftarrow \theta + \alpha \nabla \min 1 + \hat{A}$

11.Update value: $\phi \leftarrow \phi - \alpha \nabla (V_\phi(s) - V$

12.End

4. Hierarchical reinforcement learning cooperative control architecture

4.1. Synergetic mechanism of the system architecture

The system employs a dual-layer intelligent agent architecture. The Manager operates on a slower time scale to optimize State of Charge (SOC) stabilization, while the Worker functions at a faster time scale to perform tactical voltage fluctuation suppression and achieve precise power tracking. The overall system workflow is depicted in Figure 3.

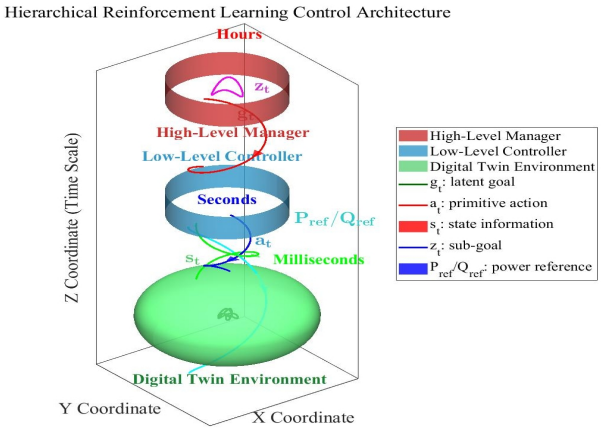

Fig. 3. The cooperative control architecture based on hierarchical reinforcement learning

In Figure 3, the layers communicate and collaborate through a potential Goal. The Manager produces an abstract

objective g_t every c time step, which serves as part of the reward function of the underlying controller, guiding its short-term behavior to align with the long-term strategy.

4.2. Multi-time scale digital twin environment enhancement modeling

To support hierarchical decision-making, the digital twin model captures both the fast dynamics (such as voltage fluctuations) and the slow dynamics (such as SOC evolution) of the system [26].

The transient process used to simulate the initial stage of the disturbance is expressed as:

$$\frac{d\mathbf{I}_{\text{line}}}{dt} = \mathbf{L}^{-1}(\mathbf{V}_{\text{bus}} - \mathbf{R}\mathbf{I}_{\text{line}} - \mathbf{V}_{\text{pcc}}) \tag{8}$$

In the formula: $\mathbf{I}_{\text{line}}$ is the line current vector (A); $\mathbf{L}$ is the line inductance matrix (H); $\mathbf{V}_{\text{bus}}$ is the bus voltage vector (V); $\mathbf{R}$ is the line resistance matrix (Ω); and $\mathbf{V}_{\text{pcc}}$ is the voltage vector at the point of common coupling (V).

The dynamic response is simulated by a first-order inertia link:

$$\frac{dP_{out}}{dt} = \frac{1}{\tau}(P_{ref} - P_{out}) \tag{9}$$

Where: P_{out} denotes the actual output active power of the PCS (W); P_{ref} denotes the active power reference command (W); τ is the time constant of the PCS (s), reflecting its response speed.

In addition to the static component, the dynamic component is added to the load model to simulate the actual disturbance more realistically. The model is expressed as:

$$T\frac{dP_{dyn}}{dt} + P_{dyn} = P_0(\frac{V}{V_0}) \tag{10}$$

In Formula 10: P_{dyn} is the dynamic active component of the load (W); T is the mechanical time constant of the motor (s); P_0 is the initial power at rated voltage (W); and α denotes the voltage index.

The model connects fast and slow time scales, and the internal control dynamics of PCS need to be more finely characterized to evaluate the tracking performance of control instructions.

5. Simulation experiment and analysis

5.1 The setup of the experimental environment

The experiments were conducted within a smart microgrid digital twin model integrating wind, solar, hydro, and storage components. The model was constructed using parameters from the Ausgrid Solar Home Electricity Dataset [27] and the Open Power System Data (OPSD) [28], which include load, wind power, and photovoltaic generation data from multiple regions. The simulation platform configuration is detailed in Tables 3 to 5.

Table 3. Hardware configuration environment

Items	Parameters	Functions
CPU	Intel Xeon Platinum 8360Y	Accelerate the simulation and training process
GPU	NVIDIA A100	Deep neural network model training and inference
RAM	512 GB DDR4 ECC	Model parameter access and exchange
DISK	Samsung PM1743 3.2TB	High-speed read and write storage log, checkpoint and data set

Table. 4 Software configuration environment

Items	Parameter/version	Functions
Operating system	Ubuntu 20.04.6 LTS	Underlying system operating environment
Programming language	Python 3.8.18	Algorithm logic, data preprocessing and analysis
Deep learning framework	PyTorch 2.1.2 + CUDA 11.8	Deployment of DRL and HRL neural network models
Power system simulation	MATLAB/Simulink R2023a	Provide a simulation environment
Interactive interface	ONNX Runtime, MATLAB Engine API	Real-time data exchange and control

Table 5. Key training hyperparameter settings [29]

Items	Parameters	Functions
Discount factor (γ)	0.99	Weigh the immediate rewards
Crop range (ϵ)	0.2	Limit policy update amplitude
Learning rate (Actor)	3e-4	Step size for updating control policy network parameters
Learning rate (Critic)	1e-3	Control the step size of the value network parameter update
Batch size (Batch Size)	1024	Number of samples used per parameter update
High-level decision-making cycle (c)	10	Interval time step for new target

5.2 Simulation results and analysis

Photovoltaic power step drop test. To evaluate the dynamic response capability of the proposed control strategy under severe fluctuations in distributed renewable energy (DRE) generation, extreme scenarios involving abrupt changes in photovoltaic output were simulated. These tests assessed the strategy's effectiveness in mitigating voltage fluctuations and maintaining the transient stability of the system. The corresponding results are presented in Table 6 and Figure 4.

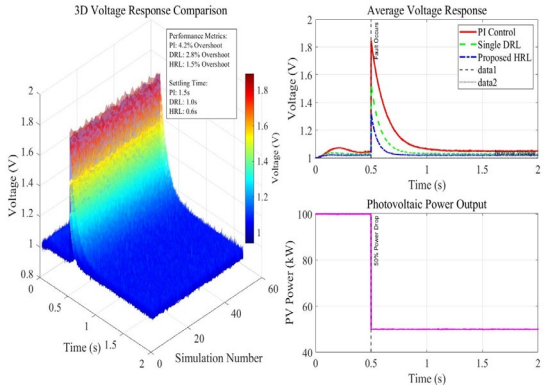

Fig. 4. Photovoltaic power step drop test

The test results demonstrate that the proposed HRL strategy exhibits superior transient performance by effectively limiting the voltage overshoot to 1.5%, which represents a reduction of 64.3% compared to the PI control and 46.4% compared to the single DRL approach. In terms of settling time, the proposed HRL requires only 0.6 seconds to restore voltage stability, which responds 60% faster than the PI control and 40% faster than the single DRL method.

Load random fluctuation test. **To evaluate the steady-state performance and stability of the control strategies under continuous random disturbances, tests were conducted by simulating continuous random variations in load power to examine the ability of each strategy to maintain voltage stability at the Point of Common Coupling (PCC). Metrics such as Root Mean Square Error (RMSE), mean deviation, and maximum deviation were computed to quantitatively assess the disturbance rejection capability and voltage control accuracy of the different control strategies [30]. The corresponding test results are presented in Table 7 and Figure 5.**

Table 6 Dynamic performance comparison under PV power step-down scenario (n=50)

Control Strategy	Overshoot (%)	Settling Time (s)	Stability Margin
PI Control	4.2	1.5	Low
Single DRL	2.8	1.0	Moderate
Proposed HRL	1.5	0.6	High

Table 7. Statistical comparison of voltage deviation under load fluctuation scenario (n=6000)

Control Strategy	Mean Deviation (V)	RMSE (V)	Maximum Deviation (V)
PI Control	1.85	2.34	8.71
Single DRL	1.12	1.42	6.53
Proposed HRL	0.78	0.90	4.89

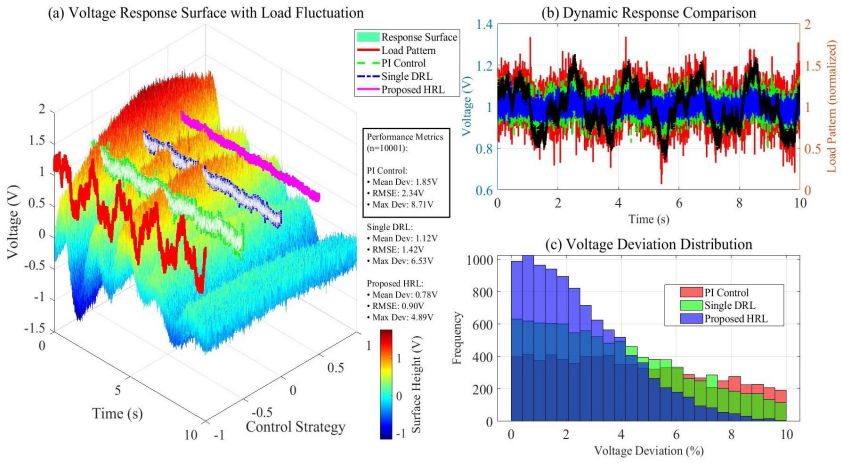

Fig. 5a. Load random fluctuation test

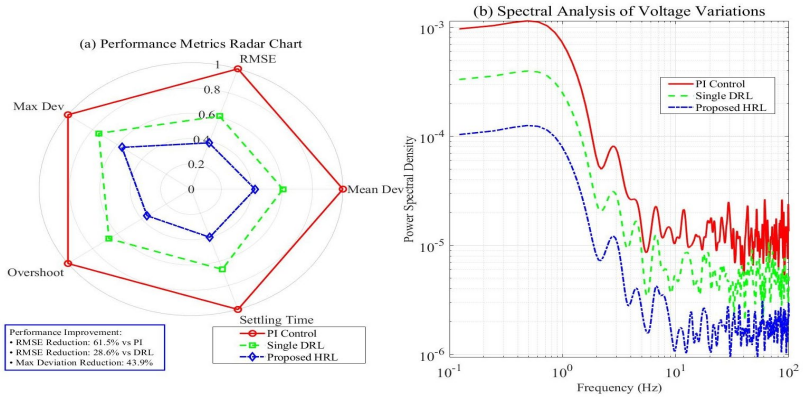

Fig. 5b. Load random fluctuation test

The test results show that the proposed HRL method reduces the voltage RMSE to $0.90\ V$, which achieves a 61.5% and 28.6% reduction compared to the PI Control and Single DRL strategies, respectively. These results corroborate that the proposed HRL can more accurately compensate for voltage deviations arising from load fluctuations, thereby providing smoother voltage support.

SOC coordination management test. The proposed HRL exhibits the capability to coordinate voltage control with energy storage management over extended time scales. To evaluate its performance under sustained disturbances, an eight-hour continuous operation scenario was simulated, during which identical stochastic sequences of renewable power and load disturbances were applied. The corresponding experimental data is presented in Table 8 and Figure 6.

Table 8. Long-term operation performance comparison (n=8 hours, data sampled per minute)

Performance Metric	Single DRL Strategy	Proposed HRL Strategy
SOC Maintenance (%)	18.5 (±5.7)	59.8 (±3.2)
Voltage RMSE (V)	3.41	0.95
Time of Regulation Loss (h)	6.5	-

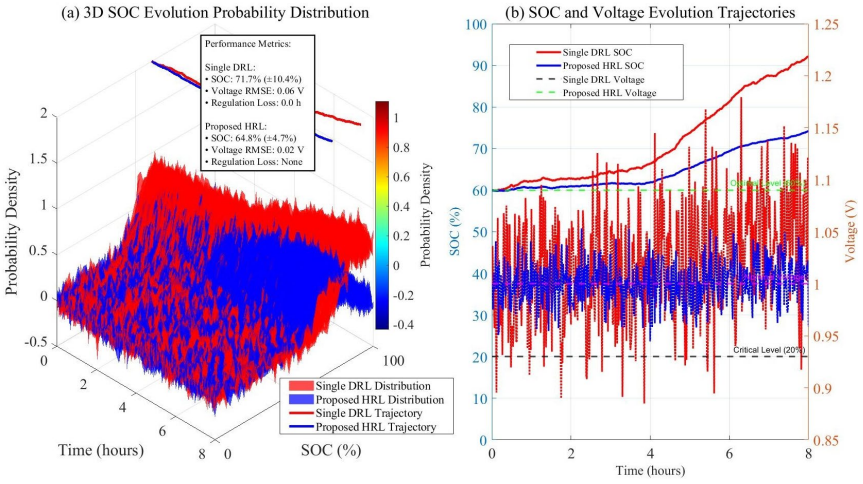

Fig. 6a. Comparison of long-term SOC evolution trajectories under different control strategies

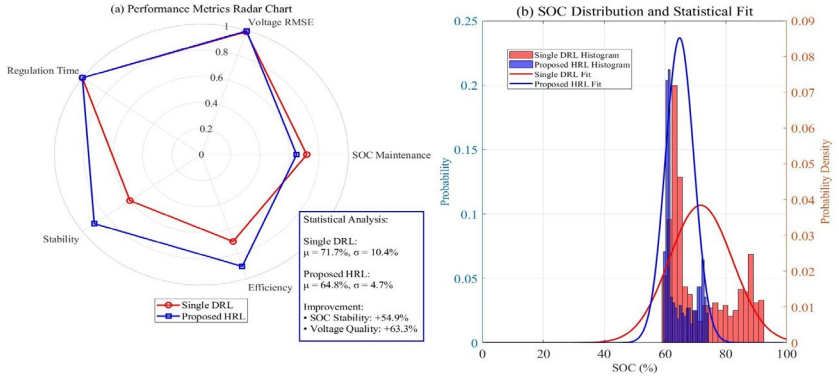

Fig. 6b. Comparison of long-term SOC evolution trajectories under different control strategie

From the results, we conclude that the Single DRL agent, whose reward function solely focuses on minimizing instantaneous voltage deviation, exhibits greedy control behavior that rapidly exhausts the energy storage capacity. In contrast, the proposed HRL framework demonstrates superior multi-objective cooperative optimization capability. Guided by latent goals generated by the Manager, the system successfully maintains the State of Charge (SOC) at a healthy average level of 59.8% (with a standard deviation of ±3.2%). Thus, it achieves an optimal balance between voltage regulation performance and energy sustainability.

5.3. Discussion

Experimental results confirm the significant potential of the proposed HRL approach in addressing multi-timescale and multi-objective control problems in microgrids. Its performance advantages primarily stem from effective functional decomposition: the Manager focuses on slow-timescale energy planning, thereby allowing the underlying controllers to concentrate on fast-timescale voltage regulation. This division of labor mitigates the curse of dimensionality and objective conflicts inherent in single-agent strategies. However, this study has limitations. For example, the ultimate performance of the algorithm partially depends on the accuracy of the digital twin model. Thus, the model inaccuracies may incur performance degradation when the strategy is deployed on physical systems.

6. Conclusion

Based on the deep integration of single-agent deep reinforcement learning (DRL) and digital twin (DT) technology, this study addresses the voltage stability issues in

smart microgrids with high penetration of distributed renewable energy (DRE) by proposing a novel coordinated control architecture for power conversion systems (PCS). A high-fidelity, multi-timescale digital twin environment was constructed. Besides, a hierarchical reinforcement learning algorithm incorporating goal-conditioned mechanisms and intrinsic reward design was developed. This framework successfully achieves effective coordination between short-term rapid and precise voltage regulation and long-term energy management. From the simulation results, we confirm that compared to conventional PI control and single-agent DRL strategies, the proposed method significantly reduces voltage deviation (with root mean square error reduced by up to 61.5%), improves dynamic response performance, and maintains state of charge (SOC) at healthy levels under power abrupt changes and stochastic load fluctuations. Therefore, the approach provides a viable solution for addressing control challenges in complex energy systems.

To mitigate potential performance degradation during real-world deployment, subsequent research will study incorporating online adaptive and transfer learning mechanisms. These enhancements are intended to improve robustness against unmodeled dynamics and model mismatch as to increase the reliability of transferring control strategies from virtual simulation to physical implementation. Further investigation will also explore the regulatory potential and economic benefits of the integrated energy systems.

References

[1] Evangeline S I, Baskaran K, Darwin S. Minimizing voltage fluctuation in stand-alone microgrid system using a Kriging-based multi-objective stochastic optimization algorithm [J]. Electrical Engineering, 2024, 106(6). DOI:10.1007/s00202-024-02497-3.

[2] Li D, Ren L, Liu F, et al. Two-time scale microgrid scheduling based on power fluctuation mitigation priority and model predictive control [J]. Energy, 2025, 324. DOI:10.1016/j.energy.2025.135760.

[3] Nandi R, Tripathy M, Gupta C P. Advanced Adaptive Virtual Impedance Based Dual Mode Inverter Controller for Power and Voltage Coordination in LV AC Microgrid [J]. IEEE Transactions on Industry Applications, 2024 (6 Pt.1): 60. DOI:10.1109/TIA.2024.3443777.

[4] Shayeghi H, Rahnama A, Bizon N. TFODn-FOPI multi-stage controller design to maintain an islanded microgrid load-frequency balance considering responsive loads support [J]. IET generation, transmission & distribution, 2023.DOI:10.1049/gtd2.12898.

[5] Liu W, Pei J, Ye Y, et al. Prescribed-performance-based adaptive fractional-order sliding mode control for ship DC microgrid [J]. Ocean Engineering, 2024, 311(Part2): 8. DOI:10.1016/j.oceaneng.2024.118885.

[6] Farrokhi E, Ghoreishy H, Ahmadiahangar R .Real-time optimal power management for a hybrid energy storage system with battery thermal consideration and DC microgrid current estimation capability [J]. Electrical Engineering, 2024, 106(4). DOI:10.1007/s00202-024-02243-9.

[7] Shi K, Yu Z, Xu P, et al. A novel adaptive control strategy for DC microgrids with additional virtual inertia [J]. Electrical Engineering, 2025, 107(8):10777-10788. DOI:10.1007/s00202-025-03060-4.

[8] Kumari P , Pan S .Enhanced load frequency control using predictive reduced order generalized active disturbance rejection control under communication delay and cyber-attack[J].Electrical Engineering, 2025, 107(3).DOI:10.1007/s00202-024-02713-0.

[9] Govindasamy S , Balapattabi S R , Kaliappan B ,et al.Energy management in microgrids using IoT considering uncertainties of renewable energy sources and electric demands: GBDT-JS approach[J].Electrical engineering, 2023(6):105.DOI:10.1007/s00202-023-01947-8.

[10] Guo J , Zhang S , Wang S .Research on nonlinear robust control strategy for active support of grid-forming photovoltaic frequency[J].The European Physical Journal Plus, 2025, 140(5):1-16.DOI:10.1140/epjp/s13360-025-06316-x.

[11] Sasidhar P R S , Gebremedhin A , Norheim I .Multi-energy microgrid design and the role of coupling components—A review[J].Renewable and Sustainable Energy Reviews, 2025, 216.DOI:10.1016/j.rser.2025.115540.

[12] Luo D , Li Z , Yan Y ,et al.Performance analysis and optimization of an annular thermoelectric generator integrated with vapor chambers[J].Energy, 2024, 307. DOI:10.1016/j.energy.2024.132565.

[13] Sahil M , Prasenjit B .Determination of microgrid stability index based on measured electrical parameters and Mamdani fuzzy inference system[J].Electrical engineering, 2024, 106(1):581-601.DOI:10.1007/s00202-023-02002-2.

[14] Chen F , Zhao Z , He X ,et al.Quantification of abnormal characteristics and flow-patterns identification in pumped storage system[J].Nonlinear Dynamics, 2024, 112(23):20813-20848. DOI:10.1007/s11071-024-10131-x.

[15] Parisio A , Rikos E , Glielmo L .A Model Predictive Control Approach to Microgrid Operation Optimization[J].IEEE Transactions on Control Systems Technology, 2014, 22(5):1813-1827. DOI:10.1109/TCST.2013.2295737.

[16] Muduli R , Jena D , Moger T .Automatic generation control of is-landed micro-grid using integral reinforcement learning-based adaptive optimal control strategy[J].Electrical Engineering, 2025, 107(3). DOI:10.1007/s00202-024-02648-6.

[17] Bhuyan M , Das D C , Barik A K . Chaotic butterfly optimization algorithm based cascaded PI-TID controller for frequency control in three area hybrid microgrid system[J].Optimal Control - Applications & Methods, 2023, 44(5). DOI:10.1002/oca.2994.

[18] Casagrande V, Ferianc M, Rodrigues M, et al. Learning-based MPC with uncertainty estimation for resilient microgrid energy management[J].IFAC PapersOnLine, 2024, 58(4):556-561. DOI:10.1016/j.ifacol.2024.07.277.

[19] Abdulwahid A H , Wang S .A new differential protection scheme for microgrid using Hilbert space based power setting and fuzzy decision processes[C]//Industrial Electronics & Applications.IEEE, 2016:6-11.DOI:10.1109/ICIEA.2016.7603542.

[20] [20] Li S , Qiu Y , Huangfu L Y .Net Power Optimization Based on Extremum Search and Model-Free Adaptive Control of PEMFC Power Generation System for High Altitude[J].IEEE transactions on transportation electrification, 2023, 9(4):5151-5164.

[21] Momen H , Jadid S .A novel microgrid formation strategy for resilience enhancement considering energy storage systems based on deep reinforcement learning[J].Journal of Energy Storage, 2024, 100.DOI:10.1016/j.est.2024.113565.

[22] Aboras K M , Alshehri M H , Megahed A I .Eel and Grouper Optimization-Based Fuzzy FOPI-TIDμ-PIDA Controller for

Frequency Management of Smart Microgrids Under the Impact of Communication Delays and Cyberattacks [J]. Mathematics (2227-7390), 2025, 13(13). DOI:10.3390/math13132040.

[23] Zheng Y , Wang Y , Meng X ,et al.Distributed Economic MPC for Synergetic Regul-ation of the Voltage of an Island DC Micro-Grid[J].Acta Automatica Sinica, 2024(3). DOI:10.1109/JAS.2023.123750.

[24] Wang C , Zhang J , Wang A ,et al.Prioritized sum-tree experience replay TD3 DRL-based online energy management of a residential microgrid[J].Applied Energy, 2024, 368.DOI:10.1016/j.apenergy.2024.123471.

[25] Musarrat M N , Fekih A ,Md. Ashib RahmanMd. Rabiul IslamKashem M. Muttaqi.An Event Triggered Sliding Mode Control-Based Fault Ride-Through Scheme to Improve the Transient Stability of Wind Energy Systems[J].IEEE Transactions on Industry Applications, 2024, 60(1 Pt.1):876-886.DOI:10.1109/TIA.2023.3328851.

[26] Cagnano A .Can Integrating SoC Management in Economic Dispatch Enhance Real-Time Operation of a Microgrid?[J].Energies (19961073), 2025, 18(7).DOI:10.3390/en18071802.

[27] Chen C , Xia R , Hu H .A novel multi-classification method for photovoltaic electricity theft behavior with low false detection rate[J].Measurement, 2025, 239(000):115461.DOI:10.1016/j.measurement.2024.115461.

[28] Wais D S , Majeed W S , Mahmood W K M .Microgrid Utilizing Solar Power System with Pulse Width Modulation[J].IOP Publishing Ltd, 2025.DOI:10.1088/1755-1315/1507/1/012011.

[29] Fadoul F F , Hassan A A , Alar R .Integrating autoencoder and decision tree models for enhanced energy consumption forecasting in microgrids: A meteorological data-driven approach in Djibouti[J].Results in Engineering, 2024, 24.DOI:10.1016/j.rineng.2024.103033.

[30] Srivastava D , Narayanan V , Singh B ,et al.A Novel Steepest Descent Least Mean Square Control for Smooth Mode Transfer of a Single-Stage SPVA-BES Hybrid Microgrid [J]. IEEE Transactions on Industry Applications. 2024, 60(5-Part1): 14. DOI:10.1109/TIA.2024.3412868.