processes

MDPI

Article

Load Frequency Control of Power Systems with an Energy Storage System Based on Safety Reinforcement Learning

Song Gao [1,*], Yudun Li [1], Xiaodi Chen [2], Zhengtang Liang [1], Enren Liu [1], Kang Liu [3] and Meng Zhang [4]

[1] State Grid Shandong Electric Power Research Institute, Ji'nan 250003, China; listu2004@126.com (Y.L.); liangzt@foxmail.com (Z.L.); liuenren0206@163.com (E.L.)
[2] School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China; xdchenin2022@163.com
[3] School of Future Technology, Xi'an Jiaotong University, Xi'an 710049, China; liukang@stu.xjtu.edu.cn
[4] School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China; tgzyyxxy@126.com
* Correspondence: gaosong3060@163.com

Abstract: Load frequency control (LFC) is a critical component in power systems that is employed to stabilize frequency fluctuations and ensure power quality. As energy storage systems (ESSs) are increasingly integrated into the grid, managing additional constraints has become more challenging. To address these challenges, this paper proposes a safety reinforcement learning-based approach that incorporates ESSs into the LFC framework. By formulating a constrained Markov decision process (CMDP), this approach overcomes the limitations of conventional Markov decision processes (MDPs) by explicitly handling system constraints. Furthermore, a long short-term memory (LSTM)-based cost prediction critic network is introduced to improve the accuracy of cost predictions, and a primal-dual deep deterministic policy gradient (PD-DDPG) algorithm is employed to solve the CMDP. Simulation results demonstrate significant improvements: a 58.2% faster settling time, a 72.5% reduction in peak frequency deviation, and a 68.2% lower mean absolute error while maintaining all operational constraints.

Keywords: load frequency control; energy storage system; safety reinforcement learning; deep deterministic policy gradient; constraint

check for updates

Academic Editor: Anthony Rossiter

Received: 28 March 2025
Revised: 25 May 2025
Accepted: 4 June 2025
Published: 16 June 2025

Citation: Gao, S.; Li, Y.; Chen, X.; Liang, Z.; Liu, E.; Liu, K.; Zhang, M. Load Frequency Control of Power Systems with an Energy Storage System Based on Safety Reinforcement Learning. *Processes* **2025**, *13*, 1897. https://doi.org/10.3390/pr13061897

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Load frequency control (LFC) serves as a fundamental mechanism for maintaining power system stability by continuously adjusting generator outputs to mitigate frequency deviations caused by load fluctuations and disturbances [1–3]. The integration of energy storage systems into modern grids has opened new avenues for enhancing frequency regulation performance. Unlike conventional generation units, ESS can provide rapid power injection/absorption to counteract frequency anomalies, offering superior dynamic response characteristics [4–6]. Among various ESS technologies, flywheel energy storage systems have emerged as particularly promising solutions for LFC applications due to their exceptional power density, ultrafast response capability, and virtually unlimited cycle life.

Recent studies have identified several technical challenges in LFC systems. Research on thermoelectric modules [7] has demonstrated the significant impact of auxiliary system losses on net energy output. Investigations of interior permanent magnet synchronous machines control systems [8] have revealed that conventional approaches often require additional sensors or observers, increasing system complexity. Furthermore, non-intrusive load monitoring studies [9,10] have highlighted the need for advanced adaptive control

capabilities to handle spatiotemporal diversity in load patterns. Significant advancements in related fields provide valuable insights for LFC optimization. Impedance analysis of grid-connected converters [11] emphasizes the critical role of accurate modeling for small-signal stability. Developments in sustainable structural batteries [12] and virtual power plant operation [13] demonstrate effective multi-objective coordination for balancing economic and security requirements. Particularly, case studies of battery-integrated renewable energy systems [14] have proven the effectiveness of bi-level optimization frameworks in resolving operational target conflicts.

However, traditional control paradigms face inherent limitations. Conventional LFC approaches are usually developed based on physical dynamic models of power systems and require high real-time computational resources. These methods often rely on linearized models and classical control techniques, which may struggle to handle the increasing complexity of modern power grids. To improve system performance, The authors of [15] proposed a unified PID tuning method for LFC, based on the two-degree-of-freedom internal model control approach, enhancing robustness and adaptability to system variations. The authors of [16] introduced a direct-indirect adaptive fuzzy control method, utilizing local frequency deviations and tie-line power deviations to dynamically adjust control strategies. The authors of [17] employed a full-order generalized state observer in a two-layer active disturbance rejection control framework, integrating equivalent input disturbance compensation to improve disturbance rejection capabilities and ensure high dynamic performance. The authors of [18] developed a robust control scheme for multi-area power systems using second-order sliding mode control combined with an extended disturbance observer, treating load variations and tie-line power deviations as lumped disturbances to enhance system stability. The authors of [19] investigated a model predictive control approach for interconnected power systems, utilizing a simplified Nordic power system model to optimize frequency control while accounting for operational constraints such as tie-line power flow limits and generation rate constraints. The authors of [20] present a comprehensive approach integrating system modeling, simulation, and optimization with advanced control techniques. Their work combines particle swarm optimization with proportional-integral-derivative controllers to enhance LFC system performance. The authors of [21] developed a novel tube-based distributed model predictive control scheme, comprising a nominal MPC component and an ancillary feedback mechanism, to achieve optimal inter-area generation coordination in multi-area power systems. Nevertheless, the growing integration of energy storage systems imposes supplementary operational constraints that are inherently difficult to capture through conventional physical modeling approaches. These emerging constraints (i) substantially elevate the complexity of LFC system design and (ii) pose significant challenges for real-time optimization processes, particularly under high renewable penetration scenarios.

To address these challenges, reinforcement learning (RL), including deep reinforcement learning (DRL), has emerged as a promising solution due to its powerful searching and learning capabilities in uncertain and complex environments. Unlike conventional model-based controllers, RL-based approaches can adaptively learn optimal control policies without requiring an explicit physical model of the system. Notable research efforts in this area have been made. The authors of [22] pioneered a correlated Q(λ) learning algorithm to optimize cooperative equilibrium control strategies in multi-area LFC systems. The authors of [23] developed a hierarchical reinforcement learning architecture featuring dual agent modules (estimator and controller), establishing an intelligent frequency regulator with enhanced uncertainty handling capabilities. The authors of [24] devised a multi-agent reinforcement learning framework employing optimized discretized joint action spaces to achieve precision LFC performance. The authors of [25] introduced a bio-inspired artificial

emotional reinforcement learning controller that synergizes mechanical logic with affective computing principles to improve adaptive control. The authors of [26] implemented a deep feature extraction paradigm using stacked denoising auto-encoders, enabling continuous action space optimization for LFC performance enhancement. The authors of [27] incorporated an attention-based critic network to selectively process environmental states, achieving reduction in training time compared to conventional methods. To address challenges including actuator malfunctions, system uncertainties, and limited communication bandwidth, the authors of [28] developed innovative event-triggered LFC strategies with reliability guarantees. In [29], a distributed observer-based framework was established for event-triggered LFC in multi-area power systems vulnerable to cyber threats. The authors of [30] proposed a model-assisted reinforcement learning approach, where an emulator network approximates power system dynamics and zeroth-order optimization enables gradient estimation for control policy generation.

The existing studies predominantly employ direct control strategies, wherein RL agents generate operational commands through real-time frequency signal processing. While these approaches have exhibited satisfactory performance in power system implementations, purely data-driven deep reinforcement learning (DRL) methods encounter fundamental limitations in precisely modeling the nonlinear dynamics of power systems during the training phase. This modeling inaccuracy may consequently yield either suboptimal control decisions or, in severe cases, system instability.

In light of these challenges, this paper proposes a safety reinforcement learning-based approach to enhance LFC performance in power systems integrated with energy storage systems. Specifically, we formulate the problem as a Constrained Markov Decision Process (CMDP), which extends the conventional MDP framework by incorporating explicit constraint handling through a cost function. This formulation enables the proposed method to effectively balance frequency regulation performance with operational constraints. Furthermore, a long short-term memory (LSTM)-based cost prediction critic network is introduced, leveraging supervised learning techniques to improve cost prediction accuracy. To solve the CMDP, we employ a primal-dual Deep Deterministic Policy Gradient (PD-DDPG) algorithm, which efficiently accommodates nonlinear constraints while optimizing frequency control performance. The main novelty of this work lies in its integrated safety reinforcement learning framework for LFC, which fundamentally advances prior approaches through two key innovations: (1) a CMDP formulation that explicitly incorporates operational constraints into the optimization framework, providing provable safety guarantees compared to existing heuristic penalty-based methods, and (2) a PD-DDPG algorithm enhanced with an LSTM-based critic network, which not only solves the CMDP through Lagrangian relaxation but also captures temporal dependencies in cost prediction—addressing critical limitations of standard DDPG implementations and feedforward critics in current literature. This dual innovation enables rigorous constraint satisfaction while achieving superior dynamic performance in renewable-rich power systems.

The rest of the paper is organized as follows. The linearized LFC model and nonlinear behaviors used throughout the paper are introduced in Section 2. Section 3 presents the framework of the proposed method. Section 4 verifies the effectiveness and advantages of the proposed method with simulation results. Finally, the conclusion of this paper is presented in Section 5.

2. Power System Frequency Response Model

This section briefly introduces the linearized LFC model and nonlinear behaviors.

2.1. Linearized LFC Model

In this paper, the microgrid includes a diesel generator and a flywheel energy storage system, with the structure shown in Figure 1.

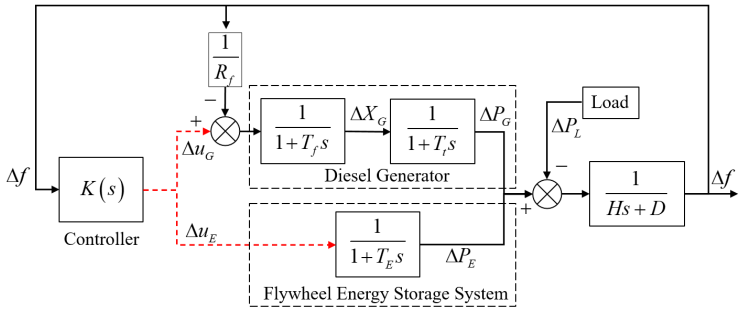

Figure 1. Schematic diagram of the linearized LFC model with a diesel generator and a flywheel energy storage system.

The whole LFC system is composed of a generator, governor, turbine, and other links involved in the secondary control loop [31]. The system dynamics can be expressed as Equations (1)–(3).

$$\Delta \dot{f} = \frac{1}{H}(\Delta P_G + \Delta P_E - \Delta P_L) - \frac{D}{H}\Delta f \tag{1}$$

$$\Delta \dot{P_G} = \frac{1}{T_t}\Delta X_G - \frac{1}{T_t}\Delta P_G \tag{2}$$

$$\Delta \dot{X_G} = \frac{1}{T_f}\Delta u_G - \frac{1}{R_f T_f}\Delta f - \frac{1}{T_f}\Delta X_G, \tag{3}$$

where ΔP_G and Δu_G represent the diesel generator's output power and control signal, respectively. ΔX_G is the output signal of the governor. T_f and T_t are the time constants of the governor and the diesel generator, respectively. R_f is the droop coefficient of the diesel generator. Δf denotes the frequency deviation of the microgrid. H represents the inertia constant. D represents the generator damping coefficient.

The flywheel energy storage system model is represented as a first-order transfer function, with its dynamic model expressed as Equation (4).

$$\Delta \dot{P_E} = \frac{1}{T_E}\Delta u_E - \frac{1}{T_E}\Delta P_E, \tag{4}$$

where ΔP_E and Δu_E represent the output power and control signal of the flywheel energy storage system, respectively. T_E is the time constant of the flywheel energy storage system.

2.2. Nonlinear Behaviors

For the sake of the practical application of the proposed method, it is necessary to model nonlinear characteristics of power systems imposed by physical dynamics. This paper considers two nonlinear power characteristics of the diesel generator: the generation rate constraint (GRC) and the power increment change constraint, as shown in Equations (5) and (6).

$$\Delta P_G = \min(\Delta P_{Gmax}, \max(0, \Delta P_G)) + \max(\Delta P_{Gmin}, \min(0, \Delta P_G)) \tag{5}$$

$$\Delta P_G = \int \min\left(\delta, \max\left(\frac{dP_g(t)}{dt}\right)\right) + max\left(-\delta, \min\left(\frac{dP_g(t)}{dt}\right)\right)dt, \tag{6}$$

where $\max(x, y)$ represents the maximum value between x and y. $\min(x, y)$ represents the minimum value between x and y. δ and $-\delta$ represent the upper and lower limits of the diesel generator's power ramp rate, respectively. ΔP_{Gmax} and ΔP_{Gmin} represent the upper and lower limits of the power increment, respectively. The power increment constraints of the flywheel energy storage system is also considered, as shown in Equation (7).

$$\Delta P_E = \min(\Delta P_{Emax}, \max(0, \Delta P_E)) + \max(\Delta P_{Emin}, \min(0, \Delta P_E)) \tag{7}$$

where ΔP_{Emax} and ΔP_{Emin} represent the upper and lower limits of the power increment of the flywheel energy storage system, respectively.

3. Main Results

3.1. Constrained Markov Decision Process

Figure 2 shows the framework of the proposed Primal-Dual DDPG. The LFC model considering the flywheel energy storage system is transformed into a CMDP. In the CMDP, if the control strategy violates the constraints, the agent will receive a negative reward. However, the size of the penalty value is difficult to define. The CMDP extends the MDP with a cost function, which can better balance future discounted rewards with control safety constraints.

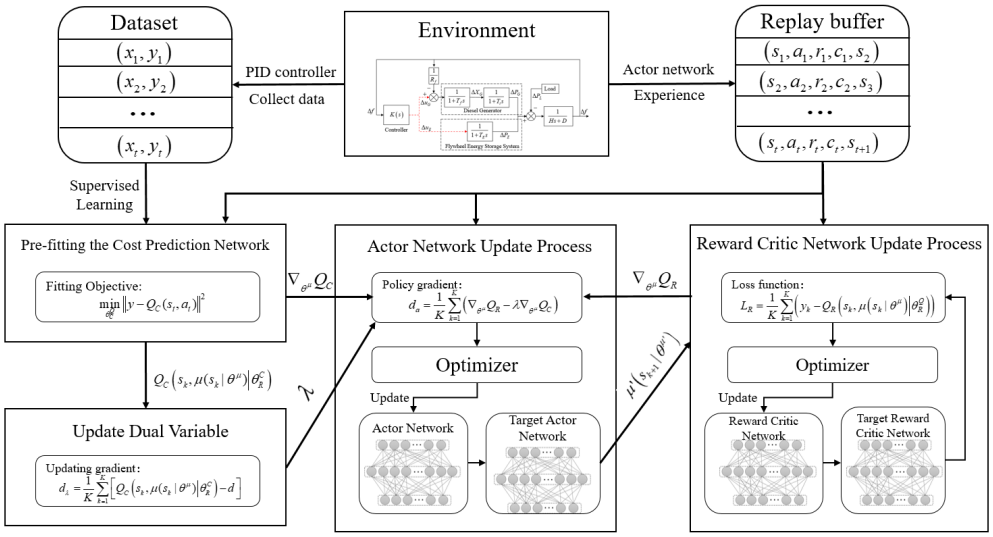

Figure 2. Framework of the proposed Primal-Dual DDPG.

In each Markov Decision Process, the agent, at each time step, observes the current state $\mathbf{s}_t$ of the power system, performs action a_t, obtains a reward value r_t and a cost value c_t, and transitions to the next state $\mathbf{s}_{t+1}$, until the total control time period is reached. The formulation is as follows. Based on the LFC problem considering the flywheel energy storage system, the CMDP for the time step is constructed as follows:

The state space of the CMDP is defined as Equation (8).

$$s_t = \left[\Delta f_t, \int \Delta f dt, \frac{d}{dt}\Delta f\right], \tag{8}$$

where $\int \Delta f dt$ and $d\Delta f/dt$ represent the integral and derivative of the frequency deviation, respectively.

The action space of the CMDP is defined as Equation (9).

$$a_t = \left[\Delta u_G^t, \Delta u_E^t\right]. \tag{9}$$

The reward function of the CMDP is defined as Equation (10).

$$r_t = \begin{cases} \alpha & |\Delta f_t| < 0.02Hz \\ \xi_1 \Delta f_t^2 & 0.02Hz \leqslant \Delta f_t < 0.05Hz \\ \xi_2 \Delta f_t^2 & 0.05Hz \leqslant \Delta f_t < 0.10Hz \\ \xi_3 \Delta f_t^2 & 0.10Hz \leqslant \Delta f_t < 0.15Hz \\ \xi_4 \Delta f_t^2 & 0.15Hz \leqslant \Delta f_t \leqslant 0.20Hz \\ \beta & \Delta f_t > 0.20Hz \end{cases}, \tag{10}$$

where α is the positive reward value, given only when the frequency deviation is less than 0.02 Hz, with $\alpha > 0$. ξ_1, ξ_2, ξ_3, and ξ_4 are the penalty coefficients corresponding to different frequency deviation intervals with $\xi_4 < \xi_3 < \xi_2 < \xi_1 < 0$. The penalty value β is applied when the frequency deviation exceeds a certain threshold, and the agent will receive a large penalty.

The cost function of the CMDP is defined as Equation (11).

$$\begin{cases} c_t = & c_G^t + c_E^t + c_\delta^t \\ c_G^t = & \max(0, \Delta P_G^t - \Delta P_{Gmax}) \\ & + \max(0, \Delta P_{Gmin} - \Delta P_G^t) \\ c_E^t = & \max(0, \Delta P_E^t - \Delta P_{Emax}) \\ & + \max(0, \Delta P_{Emin} - \Delta P_E^t) \\ c_\delta^t = & \max\left(0, {}^{dP_G^t}/_{dt} - \delta\right) \\ & + \max\left(0, -\delta - {}^{dP_G^t}/_{dt}\right) \end{cases}, \tag{11}$$

C_G^t and C_δ^t are the functions representing the violation of the power generation rate and the power increment change of the diesel generator, respectively. C_E^t is the function representing the violation of the power generation rate of the flywheel energy storage system.

The objective function of the CMDP is defined as Equation (12), which builds directly upon the standard CMDP framework established in [32,33].

$$\begin{cases} \pi^* = & \arg\max_\pi R(\pi) = \mathop{\mathbb{E}}_\pi \left[\sum_{t=0}^{T} \gamma^t r_t | s_t, a_t \right] \\ s.t. & C(\pi) = \mathop{\mathbb{E}}_\pi \left[\sum_{t=0}^{T} \gamma^t c_t | s_t, a_t \right] \leqslant d \end{cases}, \tag{12}$$

where γ is the discount factor, and $0 \leq \gamma \leq 1$. π is the policy mapping from states to actions. r_t is the reward at time t. c_t is the cost at time t. T is the total control time period. The objective is to maximize the future discounted rewards while satisfying the control safety constraints.

3.2. LSTM-Based Cost Critic Network

A cost prediction network based on LSTM is proposed. This network takes the state and action of the frequency control model as inputs to predict the control cost at the current time step, thus realizing a nonlinear mapping from state–action pairs to costs. This method can capture the complex temporal dependencies between states and actions during the system's operation, effectively improving the accuracy of cost prediction.

In this framework, a PID controller is used to control the LFC model considering the flywheel energy storage system. By collecting real-time information from the power system, a labeled dataset $D = \{(s_t, a_t, c_t)\}$ is constructed, where the dataset includes the inputs and

outputs of the PID controller, i.e., s_t and a_t, respectively. Furthermore, the labeled dataset is used to train the LSTM network, thereby establishing a mapping from state and action to cost via supervised learning. During actual control, the cost calculated using Equation (11) is also used as the label for training data. By utilizing the dynamic response data of the flywheel energy storage system, an effective data pair for constructing the LSTM model is formed.

The cost prediction network based on the LSTM architecture is shown in Equation (13). The cost prediction network receives input x_t at time step t and updates the current state through hidden states h_t and memory cells c_{tLSTM}. The update equation for the cost prediction network's state is shown as Equation (13).

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{c}_{tLSTM} = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ c_{tLSTM} = f_t \odot c_{tLSTM} + i_t \odot \tilde{c}_{tLSTM} \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \cdot tanh(c_{tLSTM}) \end{cases}, \tag{13}$$

where f_t, i_t, and o_t are the forget, input, and output gates, respectively, $\odot$ represents element-wise multiplication, σ is the activation function, W_f, W_i, W_o, and W_c are weight matrices, and b_f, b_i, b_o, and b_c are bias terms.

Once the updated hidden state h_t is obtained, the output of the cost prediction network is mapped to the predicted cost via a fully connected layer, shown as Equation (14).

$$Q_c(s_t, a_t) = h^{(n)}\left[h^{(n-1)}\left[\cdots h^{(1)}(h_t)\right]\right] \tag{14}$$

where $h^{(n)}$ is the activation function of the output layer. $Q_c(s_t, a_t)$ is the predicted cost.

The cost prediction network is trained using supervised learning. The training of the cost prediction network aims to minimize the difference between the actual control cost and the cost predicted by the network at each time step, as shown in Equation (15):

$$\min_{\theta_c^Q} \|y - Q_c(s_t, a_t)\|^2 \tag{15}$$

where θ_c^Q represents the parameter of the cost prediction network. y is the label of the dataset.

3.3. The Proposed Primal-Dual Ddpg

A method based on Primal-Dual DDPG is proposed for solving CMDP. By introducing a Lagrange multiplier, the inequality constraint problem in the objective function is transformed into an unconstrained problem. Additionally, by utilizing a cost prediction neural network, the efficiency of the safe reinforcement learning algorithm is improved, enabling the update of the actor network.

The original CMDP problem is shown in Equation (12). To solve this, we introduce a Lagrange multiplier λ and reformulate the problem as an unconstrained min–max optimization shown in Equation (16).

$$\mathcal{L}(\pi, \lambda) = R(\pi) - \lambda(C(\pi) - d) \tag{16}$$

The primal-dual optimization then becomes Equation (17):

$$\max_{\pi} \min_{\lambda \geq 0} \mathcal{L}(\pi, \lambda) \tag{17}$$

Based on PD-DDPG [34], the Lagrange multiplier is updated to penalize constraint violations, shown in Equation (18).

$$\lambda \leftarrow \max(0, \lambda - \eta_\lambda (C(\pi) - d)) \tag{18}$$

where η_λ is the learning rate for λ. In our method, $C(\pi)$ is replaced by the LSTM-based cost critic network $Q_c(s_t, a_t)$. The policy π is trained to maximize $\mathcal{L}(\pi, \lambda)$, effectively balancing reward maximization and constraint satisfaction.

A deep reinforcement learning controller is constructed, consisting of the actor network $\mu(a|\theta^\mu)$, the reward critic network $Q_R(s, a|\theta_R^Q)$, the target actor network $\mu'(a|\theta^{\mu'})$, and the target reward critic network $Q'_R(s, a|\theta_R^{Q'})$. The parameters of the actor network θ^μ and the reward critic network θ_R^Q are initialized, and the parameters of both networks are subsequently assigned to their respective target networks.

The interaction experience pool and dual variable λ are initialized. Interaction experiences between the actor network and the power system are obtained through simulation experiments to populate the experience pool.

The parameters of the actor network, reward critic network, target actor network, and target reward critic network are optimized based on the interaction experience pool. Specifically, the optimization process includes the following steps based on Equations (19)–(27).

1. Target reward critic value calculation: K interaction experiences are sampled from the experience pool $\{(s_k, a_k, r_k, c_k, s_{k+1})\}_{k=1}^K$. The target reward critic value $Q_{\text{target},k}$ is computed using the equation:

$$Q_{\text{target},k} = r_k + \gamma Q'_R\left(s_{k+1}, \mu'\left(s_{k+1}|\theta^{\mu'}\right)\middle|\theta_R^{Q'}\right) \tag{19}$$

   where $\theta^{\mu'}$ and $\theta_R^{Q'}$ denote the parameters of the target actor network and the target reward critic network, respectively. γ is the discount factor. r_k is the immediate reward at step k. s_{k+1} is the next state from experience tuple $\{(s_k, a_k, r_k, c_k, s_{k+1})$.
2. Updating reward critic network: The parameters of the reward critic network are updated by minimizing the loss function $\mathcal{L}_R$:

$$\mathcal{L}_R = \frac{1}{K}\sum_{k=1}^{K}\left(Q_{\text{target},k} - Q_R\left(s_k, \mu(s_k|\theta^\mu)\middle|\theta_R^Q\right)\right) \tag{20}$$

3. Action gradient calculation: Based on Equation (18), the action gradient d_a is computed using

$$d_a = \frac{1}{K}\sum_{k=1}^{K}(\nabla_{\theta^\mu} Q_R - \lambda \nabla_{\theta^\mu} Q_C) \tag{21}$$

   where $\nabla_{\theta^\mu} Q_R$ and $\nabla_{\theta^\mu} Q_C$ are computed based on chain rule, shown in Equations (22) and (23).

$$\nabla_{\theta^\mu} Q_R = \nabla_a Q_R\left(s, \mu(s|\theta^\mu)\middle|\theta_R^Q\right)\nabla_{\theta^\mu}\mu(s|\theta^\mu)\Big|_{s=s_k} \tag{22}$$

$$\nabla_{\theta^\mu} Q_C = \nabla_a Q_C\left(s, \mu(s|\theta^\mu)\middle|\theta_C^Q\right)\nabla_{\theta^\mu}\mu(s|\theta^\mu)\Big|_{s=s_k} \tag{23}$$

where $\nabla_a Q_R\left(s, \mu(s|\theta^\mu)|\theta_R^Q\right)$, $\nabla_a Q_C\left(s, \mu(s|\theta^\mu)|\theta_C^Q\right)$, and $\nabla_{\theta^\mu}\mu(s|\theta^\mu)$ are derived from the reward critic network, cost prediction network, and actor network, respectively, using the backpropagation algorithm. The parameters of the actor network are updated using gradient descent:

$$\theta^\mu \leftarrow \theta^\mu - \eta_a \cdot d_a \tag{24}$$

where η_a is the learning rate of the actor network.

4. Dual variable gradient calculation: Based on Equation (18), the gradient of the dual variable λ is calculated as Equation (25).

$$d_\lambda = \frac{1}{K}\sum_{k=1}^{K}\left[Q_C(s_k, \mu(s_k|\theta^\mu)|\theta_C^Q) - d\right] \tag{25}$$

The dual variable are updated via gradient descent shown in Equation (26).

$$\lambda \leftarrow \max(0, \lambda - \eta_\lambda d_\lambda) \tag{26}$$

where η_λ is the learning rate of the dual variable.

5. Soft update of target networks: The target actor network and target reward critic network are updated using the soft update rule:

$$\begin{cases} \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \\ \theta_R^{Q'} \leftarrow \tau\theta_R^Q + (1-\tau)\theta_R^{Q'} \end{cases} \tag{27}$$

where τ is a preset soft update parameter.

4. Case Study

The proposed PD-DDPG LFC method is evaluated on a power system with parameters detailed in Table 1, which also outlines the system constraints. For the controller updating in simulations, we set the size of the memory replay buffer, learning rate η, and constant τ for updating the target network h as 8000, 0.00005, and 0.1, respectively. DDPG [35] is used as a comparison algorithm. DDPG applies different fixed penalty coefficients to penalize actions that violate constraints, with the penalty factor defined as

$$r_p = r_t - \rho c_t \tag{28}$$

where ρ is the penalty coefficient, set to 0.001 and 0.005, respectively. A PID controller, whose data builds the LFC database, is also used for comparison, and its gains are tuned manually. The memory replay buffer is initialized by the LFC database. Figure 3 shows that the final converged loss of the LSTM-based network is significantly lower than that of the CNN-based network. Figure 4 shows the prediction cost comparison between CNN-based and LSTM-based cost critic networks, demonstrating that the LSTM-based network achieves better prediction cost performance than the CNN-based network.

Table 1. Parameters of the power system.

Parameter	Value	Parameter	Value	Parameter	Value
ΔP_{Gmax} (p.u.)	0.03	T_f (s)	10	H (p.u./Hz)	14.22
ΔP_{Gmin} (p.u.)	−0.03	T_t (s)	0.10	D (p.u./Hz)	0
ΔP_{Emax} (p.u.)	0.025	δ (p.u./s)	0.0017 [36]	R (Hz/p.u.)	0.05
ΔP_{Emin} (p.u.)	−0.025	T_E (s)	1		

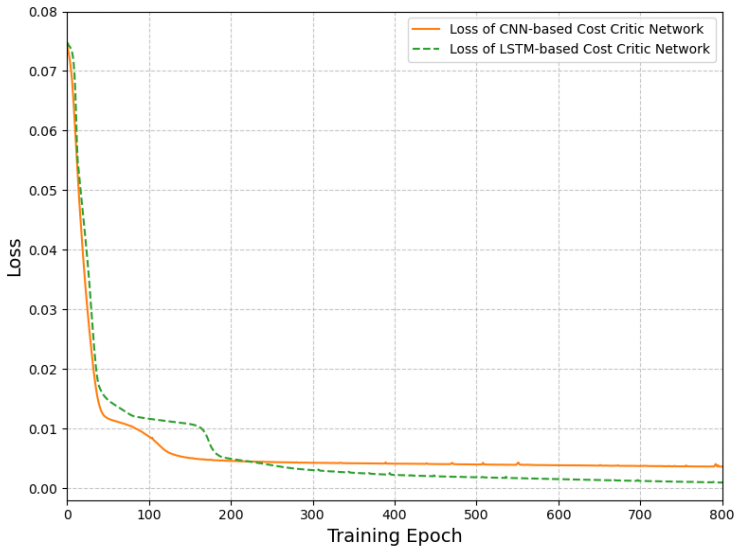

Figure 3. Training loss comparison of CNN-based vs. LSTM-based cost critic network.

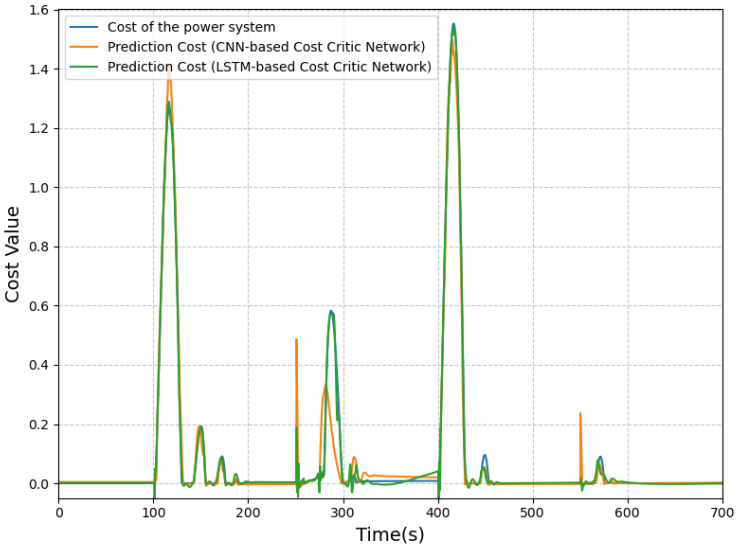

Figure 4. Prediction cost comparison of CNN-based vs. LSTM-based cost critic network. Note: the LSTM-based network demonstrates superior prediction accuracy for system operational costs compared to the CNN-based approach.

Step disturbances are applied at the following times: a 0.04 $p.u.$ disturbance at $t = 100$ s, a -0.06 $p.u.$ disturbance at $t = 250$ s, a 0.05 $p.u.$ disturbance at $t = 400$ s, and a -0.03 $p.u.$ disturbance at $t = 550$ s. Figure 5 shows the frequency deviation of the power system. It can be observed that, although the DDPG method with a penalty factor ($\rho = 0.005$) performs better than the proposed PD-DDPG method when handling positive step disturbances, it performs worse than even the PID controller when facing negative step disturbances. Overall, the proposed PD-DDPG method provides the best control performance, with a smaller overshoot in the frequency deviation and a faster recovery speed, demonstrating its superior ability to regulate frequency. As shown in Table 2, the proposed PD-DDPG achieves superior overall performance despite a 13.8% longer settling time (55.02 s) versus DDPG (ρ = 0.001) (48.33 s), delivering 41.4% lower peak deviation (0.0061 p.u.) and 30% reduced MAE (0.0007 p.u.) while maintaining 86.3% faster response than conventional PID. This optimal trade-off ensures enhanced transient stability and robustness, where minimizing overshoots proves more critical than absolute settling time in practical power

systems. Figures 6 and 7 show the power output of the diesel generator and the flywheel energy storage system, respectively. It can be observed that, due to different constraint conditions, the power output response curves of the diesel generator and the flywheel energy storage system exhibit distinct characteristics. Under the PD-DDPG control strategy, both the diesel generator and the flywheel energy storage system better account for these constraints during the update process. The proposed method coordinately reduces the power output fluctuations of both diesel generator and flywheel energy storage system, significantly optimizing the amplitude of system frequency deviation fluctuations and effectively improving frequency regulation performance.

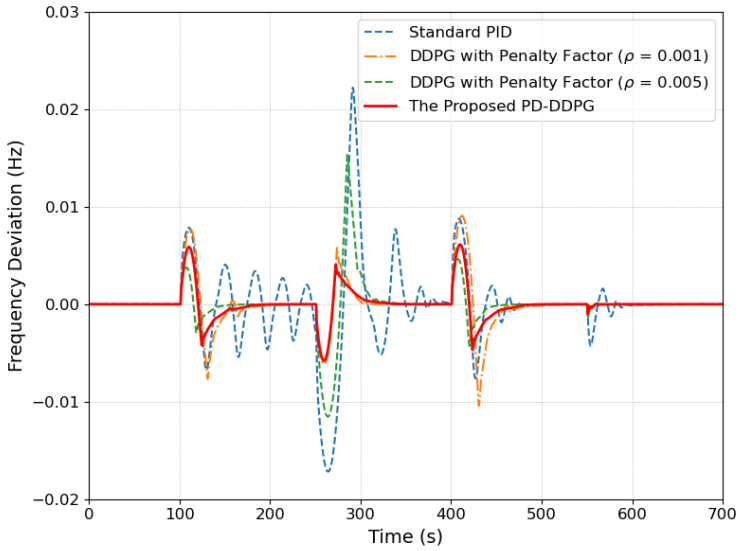

Figure 5. Frequency deviation using different methods tested on the power system.

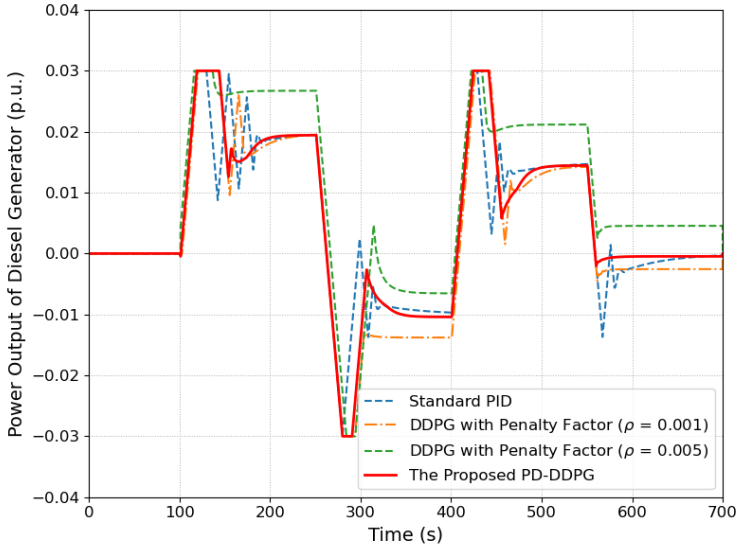

Figure 6. Power output of the diesel generator using different methods. Note: The slope demonstrates full compliance with GRC, while maintaining strict output boundaries of [−0.03, 0.03] p.u. throughout operation.

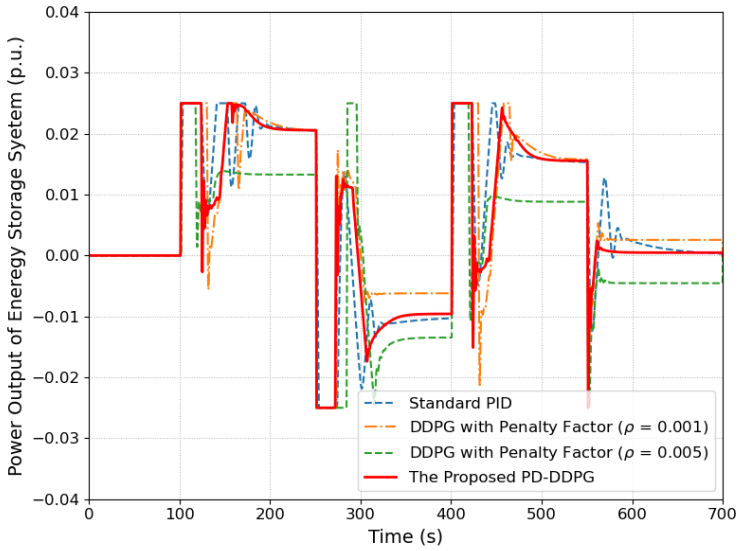

Figure 7. Power output of the flywheel energy storage system using different methods. Note: The operational constraints strictly limit FESS power output within $[-0.025, 0.025]$p.u.

Table 2. Performance metrics of different methods.

Method	Settling Time (s)	Peak (p.u.)	Mean Absolute Error (MAE) (p.u.)
Standard PID	131.51	0.0222	0.0022
DDPG [35] with $\rho = 0.001$	48.33	0.0104	0.0010
DDPG [35] with $\rho = 0.005$	70.50	0.0153	0.0008
Proposed PD-DDPG	55.02	0.0061	0.0007

5. Conclusions

This paper proposes a safety reinforcement learning framework for load frequency control in modern power systems with energy storage integration, featuring a constrained Markov decision process formulation solved through a novel primal-dual deep deterministic policy gradient algorithm enhanced by an LSTM-based critic network. The proposed approach demonstrates superior constraint-handling capability and frequency regulation performance compared to conventional methods, particularly in renewable-rich power systems, while maintaining strict operational limits, thereby providing an effective solution for next-generation smart grid applications.

Author Contributions: Conceptualization, S.G. and Y.L.; methodology, S.G.; software, S.G. and X.C.; validation, Y.L. and X.C.; formal analysis, S.G.; investigation, Z.L. and E.L.; resources, Z.L. and E.L.; data curation, X.C.; writing—original draft preparation, S.G.; writing—review and editing, X.C.; visualization, X.C. and K.L.; supervision, Y.L.; project administration, S.G. and M.Z.; funding acquisition, Y.L., K.L. and M.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Science and Technology Project of the State Grid Shandong Electric Power Research Institute, "Research and Application of Collaborative Control Technology for Multi-type Energy Storage and Temporal Complementarity" (No. 520626240009).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Song Gao, Yudun Li, Zhengtang Liang and Enren Liu were employed by the State Grid Shandong Electric Power Research Institute. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Zhao, C.; Topcu, U.; Li, N.; Low, S. Design and stability of load-side primary frequency control in power systems. *IEEE Trans. Autom. Control* **2014**, *59*, 1177–1189. [CrossRef]
2. Shi, T.; Wang, C.; Chen, Z. The multi-point cooperative control strategy for electrode boilers supporting grid frequency regulation. *Processes* **2025**, *13*, 785. [CrossRef]
3. Singh, V.P.; Kishor, N.; Samuel, P. Load frequency control with communication topology changes in smart grid. *IEEE Trans. Ind. Inform.* **2016**, *12*, 1943–1952. [CrossRef]
4. El-Rifaie, A.M. A novel lyrebird optimization algorithm for enhanced generation rate-constrained load frequency control in multi-area power systems with proportional integral derivative controllers. *Processes* **2025**, *13*, 949. [CrossRef]
5. Huynh, V.V.; Minh, B.L.N.; Amaefule, E.N.; Tran, A.-T.; Tran, P.T. Highly robust observer sliding mode based frequency control for multi area power systems with renewable power plants. *Electronics* **2021**, *10*, 274. [CrossRef]
6. Zhang, C.-K.; Jiang, L.; Wu, Q.; He, Y.; Wu, M. Delay-dependent robust load frequency control for time delay power systems. *IEEE Trans. Power Syst.* **2013**, *28*, 2192–2201. [CrossRef]
7. Miao, Z.; Meng, X.; Li, X.; Liang, B.; Watanabe, H. Enhancement of net output power of thermoelectric modules with a novel air-water combination. *Appl. Therm. Eng.* **2025**, *258*, 124745. [CrossRef]
8. Ding, S.; Liu, C.; Fan, Z.; Hang, J. Lumped parameter adaptation-based automatic mtpa control for ipmsm drives by using stator current impulse response. *IEEE Trans. Energy Convers.* **2025**.
9. Lin, L.; Liu, J.; Huang, N.; Li, S.; Zhang, Y. Multiscale spatio-temporal feature fusion based non-intrusive appliance load monitoring for multiple industrial industries. *Appl. Soft Comput.* **2024**, *167*, 112445. [CrossRef]
10. Lin, L.; Ma, X.; Chen, C.; Xu, J.; Huang, N. Imbalanced industrial load identification based on optimized catboost with entropy features. *J. Electr. Eng. Technol.* **2024**, *19*, 4817–4832. [CrossRef]
11. Rong, Q.; Hu, P.; Yu, Y.; Wang, D.; Cao, Y.; Xin, H. Virtual external perturbance-based impedance measurement of grid-connected converter. *IEEE Trans. Ind. Electron.* **2024**, *72*, 2644–2654.
12. Kusekar, S.K.; Pirani, M.; Birajdar, V.D.; Borkar, T.; Farahani, S. Toward the progression of sustainable structural batteries: State-of-the-art review. *SAE Int. J. Sustain. Transp. Energy Environ. Policy* **2025**, *5*, 283–308. [CrossRef]
13. Akbari, E.; Naghibi, A.F.; Veisi, M.; Shahparnia, A.; Pirouzi, S. Multi-objective economic operation of smart distribution network with renewable-flexible virtual power plants considering voltage security index. *Sci. Rep.* **2024**, *14*, 19136. [CrossRef] [PubMed]
14. Navesi, R.B.; Jadidoleslam, M.; Moradi-Shahrbabak, Z.; Naghibi, A.F. Capability of battery-based integrated renewable energy systems in the energy management and flexibility regulation of smart distribution networks considering energy and flexibility markets. *J. Energy Storage* **2024**, *98*, 113007. [CrossRef]
15. Tan, W. Unified tuning of pid load frequency controller for power systems via imc. *IEEE Trans. Power Syst.* **2009**, *25*, 341–350. [CrossRef]
16. Yousef, H.A.; Khalfan, A.-K.; Albadi, M.H.; Hosseinzadeh, N. Load frequency control of a multi-area power system: An adaptive fuzzy logic approach. *IEEE Trans. Power Syst.* **2014**, *29*, 1822–1830. [CrossRef]
17. Liu, F.; Li, Y.; Cao, Y.; She, J.; Wu, M. A two-layer active disturbance rejection controller design for load frequency control of interconnected power system. *IEEE Trans. Power Syst.* **2015**, *31*, 3320–3321. [CrossRef]
18. Liao, K.; Xu, Y. A robust load frequency control scheme for power systems based on second-order sliding mode and extended disturbance observer. *IEEE Trans. Ind. Inform.* **2017**, *14*, 3076–3086. [CrossRef]
19. Ersdal, A.M.; Imsland, L.; Uhlen, K. Model predictive load-frequency control. *IEEE Trans. Power Syst.* **2015**, *31*, 777–785. [CrossRef]
20. Ogar, V.N.; Hussain, S.; Gamage, K.A. Load frequency control using the particle swarm optimisation algorithm and pid controller for effective monitoring of transmission line. *Energies* **2023**, *16*, 5748. [CrossRef]
21. Liu, X.; Wang, C.; Kong, X.; Zhang, Y.; Wang, W.; Lee, K.Y. Tube-based distributed mpc for load frequency control of power system with high wind power penetration. *IEEE Trans. Power Syst.* **2023**, *39*, 3118–3129. [CrossRef]
22. Yu, T.; Wang, H.; Zhou, B.; Chan, K.W.; Tang, J. Multi-agent correlated equilibrium q (λ) learning for coordinated smart generation control of interconnected power grids. *IEEE Trans. Power Syst.* **2014**, *30*, 1669–1679. [CrossRef]
23. Singh, V.P.; Kishor, N.; Samuel, P. Distributed multi-agent system-based load frequency control for multi-area power system in smart grid. *IEEE Trans. Ind. Electron.* **2017**, *64*, 5151–5160. [CrossRef]
24. Daneshfar, F. Intelligent load-frequency control in a deregulated environment: Continuous-valued input, extended classifier system approach. *IET Gener. Transm. Distrib.* **2013**, *7*, 551–559. [CrossRef]

25. Yin, L.; Yu, T.; Zhou, L.; Huang, L.; Zhang, X.; Zheng, B. Artificial emotional reinforcement learning for automatic generation control of large-scale interconnected power grids. *IET Gener. Transm. Distrib.* **2017**, *11*, 2305–2313. [CrossRef]
26. Yan, Z.; Xu, Y. Data-driven load frequency control for stochastic power systems: A deep reinforcement learning method with continuous action search. *IEEE Trans. Power Syst.* **2018**, *34*, 1653–1656. [CrossRef]
27. Yang, F.; Huang, D.; Li, D.; Lin, S.; Muyeen, S.; Zhai, H. Data-driven load frequency control based on multi-agent reinforcement learning with attention mechanism. *IEEE Trans. Power Syst.* **2022**, *38*, 5560–5569. [CrossRef]
28. Zhang, M.; Dong, S.; Wu, Z.-G.; Chen, G.; Guan, X. Reliable event-triggered load frequency control of uncertain multiarea power systems with actuator failures. *IEEE Trans. Autom. Sci. Eng.* **2022**, *20*, 2516–2526. [CrossRef]
29. Zhang, M.; Dong, S.; Shi, P.; Chen, G.; Guan, X. Distributed observer-based event-triggered load frequency control of multiarea power systems under cyber attacks. *IEEE Trans. Autom. Sci. Eng.* **2022**, *20*, 2435–2444. [CrossRef]
30. Chen, X.; Zhang, M.; Wu, Z.; Wu, L.; Guan, X. Model-free load frequency control of nonlinear power systems based on deep reinforcement learning. *IEEE Trans. Ind. Inform.* **2024**, *20*, 6825–6833. [CrossRef]
31. Bevrani, H. *Robust Power System Frequency Control*; Springer: Berlin/Heidelberg, Germany, 2014; Volume 4.
32. Wachi, A.; Sui, Y. Safe reinforcement learning in constrained markov decision processes. In Proceedings of the International Conference on Machine Learning PMLR, Virtual, 13–18 July 2020; pp. 9797–9806.
33. Altman, E. *Constrained Markov Decision Processes*; Routledge: London, UK, 2021.
34. Liang, Q.; Que, F.; Modiano, E. Accelerated primal-dual policy optimization for safe reinforcement learning. *arXiv* **2018**, arXiv:1802.06480.
35. Liu, G.-X.; Liu, Z.-W.; Wei, G.-X. Model-free load frequency control based on multi-agent deep reinforcement learning. In Proceedings of the 2021 IEEE International Conference on Unmanned Systems, Beijing, China, 15–17 October 2021; pp. 815–819.
36. Morsali, J.; Zare, K.; Hagh, M.T. Appropriate generation rate constraint (grc) modeling method for reheat thermal units to obtain optimal load frequency controller (LFC). In Proceedings of the 5th Conference on Thermal Power Plants (CTPP), Tehran, Iran, 10–11 June 2014; pp. 29–34.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.