

Real-Time Demand Response Dispatch in Smart Grids: Balancing Economic Efficiency and Social Fairness via Advanced Data-Driven Approaches

Liyuan Liu¹, Junxiang Li^{1,2*}, Wenjie Chen¹

¹Business School, University of Shanghai for Science and Technology, Shanghai, China

²School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai, China

Email: Brucelau0314@163.com, *lijx@usst.edu.cn

How to cite this paper: Liu, L.Y., Li, J.X. and Chen, W.J. (2026) Real-Time Demand Response Dispatch in Smart Grids: Balancing Economic Efficiency and Social Fairness via Advanced Data-Driven Approaches. *Journal of Power and Energy Engineering*, **14**, 97-112.
<https://doi.org/10.4236/jpee.2026.144006>

Received: April 5, 2026

Accepted: April 24, 2026

Published: April 27, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Decarbonizing modern power distribution networks requires a fundamental shift from generation-centric control to bi-directional, user-participatory mechanisms. While existing demand response (DR) frameworks prioritize total operational economy, they often neglect the disparate financial impacts on heterogeneous stakeholders, leading to potential social resistance. To address this, this study develops a multi-physics aware dispatch framework that reconciles economic objectives with cross-sectoral equity. Unlike conventional data-driven models, our approach embeds a LinDistFlow-based physical safety layer and a depth-of-discharge (DoD) battery degradation model directly into the learning environment to ensure hardware stewardship and grid stability. We employ a Soft Actor-Critic (SAC) agent optimized with Jain's Fairness Index to dynamically allocate transition costs among residential, commercial, and industrial sectors. Simulation results on an IEEE 33-bus system demonstrate that the proposed strategy achieves zero voltage violations and a significant leap in social harmony, yielding a robust 92% fairness index with a negligible 3.2% sacrifice in system-wide efficiency.

Keywords

Demand Response, Deep Reinforcement Learning, Social Fairness, Smart Grid, Low-Carbon Economic Dispatch, Multi-Objective Optimization

1. Introduction

As active prosumers increasingly dominate low-carbon electricity markets, the physical boundaries of distribution networks—particularly nodal voltage limits and thermal line capacities—have emerged as the primary bottlenecks for large-

scale demand response (DR) execution [1]. Traditional economic dispatch frameworks, which predominantly optimize for system-wide cost reduction or emission targets, often operate under idealized network assumptions [2] [3]. This oversight not only risks severe reverse power flows and equipment degradation during peak flexibility events, but also disproportionately penalizes industrial stakeholders whose rigid physical production constraints restrict their responsiveness [4] [5].

Unlike previous studies focusing solely on algorithmic convergence, modern grid dispatch must evolve to deeply integrate multi-physics constraints with cross-sectoral social equity [6]. The inherent heterogeneity across residential, commercial, and industrial sectors means that a “one-size-fits-all” pricing signal often leads to a “Green Penalty,” where rigid industrial loads subsidize the flexibility dividends of other sectors [7]. To technically resolve this socio-technical tension, this paper proposes an advanced Soft Actor-Critic (SAC) framework [8]. By embedding a LinDistFlow-based physical safety layer and a physics-based battery aging model directly into the learning environment, our approach ensures hardware stewardship and grid stability while navigating the complex Pareto frontier of smart grid operations [9].

As global decarbonization goals intensify, emission-based DR and low-carbon optimization have emerged as key pathways to align economic dispatch with environmental targets [10] [11]. In integrated electricity and carbon markets, prosumer behavior is driven not only by real-time electricity prices but also by fluctuating carbon trading signals [12]. However, designing efficient and socially equitable dispatch strategies faces critical challenges. A primary hurdle lies in the significant heterogeneity across different sectors [13] [14]. While residential users possess flexible loads like HVAC and electric vehicles (EVs), their participation is constrained by the need to maintain basic living comfort. Commercial sectors exhibit high baseloads but have specific operating windows, whereas industrial users, though possessing significant demand, often face rigid production schedules and high economic risks from supply interruptions [13] [15].

Furthermore, modern grid dispatch must respect stringent physical and technical constraints [16]. This includes maintaining distribution network power flow security (e.g., voltage limits) and accounting for the non-linear degradation costs of energy storage systems during frequent cycling [8] [17]. Traditional optimization frameworks, such as bi-level programming or Stackelberg games, often struggle to handle these high-dimensional, non-convex constraints in real-time [9] [11]. While deep reinforcement learning (DRL) has shown promise in managing such complexities [18]-[22], most existing research focuses on total system efficiency, often overlooking the “fairness gap” among different user sectors [23] [24]. In a low-carbon transition, a purely efficiency-driven strategy may inadvertently impose a disproportionate financial burden on industrial production or over-exploit residential flexibility, leading to sectoral inequities that undermine the social sustainability of DR programs.

To address these challenges, this paper proposes an advanced DRL-based

framework for low-carbon economic dispatch that explicitly balances economic efficiency with sectoral fairness. By replacing the abstract mathematical equity used in previous studies with a tangible cross-sector fairness metric, this study ensures a balanced distribution of transition costs and benefits among residential, commercial, and industrial stakeholders. The main contributions of this work are as follows:

1) **Multi-Physics Constrained Environment for Hardware Stewardship:** We formulate a dispatch environment that transcends abstract optimization by integrating non-linear AC power flow constraints and a physics-based battery aging model. This ensures that the reinforcement learning agent respects the physical limits of the distribution infrastructure while preventing the over-exploitation of energy storage assets.

2) **Sectoral Fairness-Aware Reward Mechanism:** We introduce a novel multi-objective reward structure grounded in the “Social Contract” theory. By utilizing Jain’s Fairness Index, the framework explicitly quantifies and mitigates the “Green Penalty” traditionally imposed on rigid industrial loads, ensuring that the economic dividends of low-carbon transitions are shared proportionally across the residential-commercial-industrial nexus.

3) **Validation of Efficiency-Equity Pareto Equilibrium:** Through extensive simulations, we provide a quantitative map of the Pareto frontier between aggregate social welfare and distributional equity. We demonstrate that through sophisticated agent design, grid operators can achieve millisecond-level responsiveness and guaranteed physical security without requiring additional infrastructure investments.

2. Problem Formulation

2.1. The Grid as a Socio-Technical Ecosystem: A Multi-Sector Perspective

The modern smart grid is no longer a simple unidirectional energy delivery system but has evolved into a complex socio-technical ecosystem where technical stability, economic efficiency, and social equity are deeply intertwined. In our framework, the Grid Dispatch Center acts as a central coordinator, tasked with reconciling the stochastic nature of renewable energy with the diverse objectives of three macro-sectors: Residential, Commercial, and Industrial.

The interaction within this ecosystem is modeled as a dynamic feedback loop. The dispatch center broadcasts price and carbon signals, and in response, each sector optimizes its internal resources to achieve a balance between economic savings and operational constraints. This “Social Contract” ensures that the burden of grid regulation and the benefits of low-carbon transitions are shared across society.

Figure 1 illustrates the holistic socio-technical architecture of the proposed multi-objective demand response market. At the top layer, the power supply side comprises conventional thermal generators and renewable energy sources, which

are collectively governed by a Tiered Carbon Cap-and-Trade mechanism to enforce environmental stewardship.

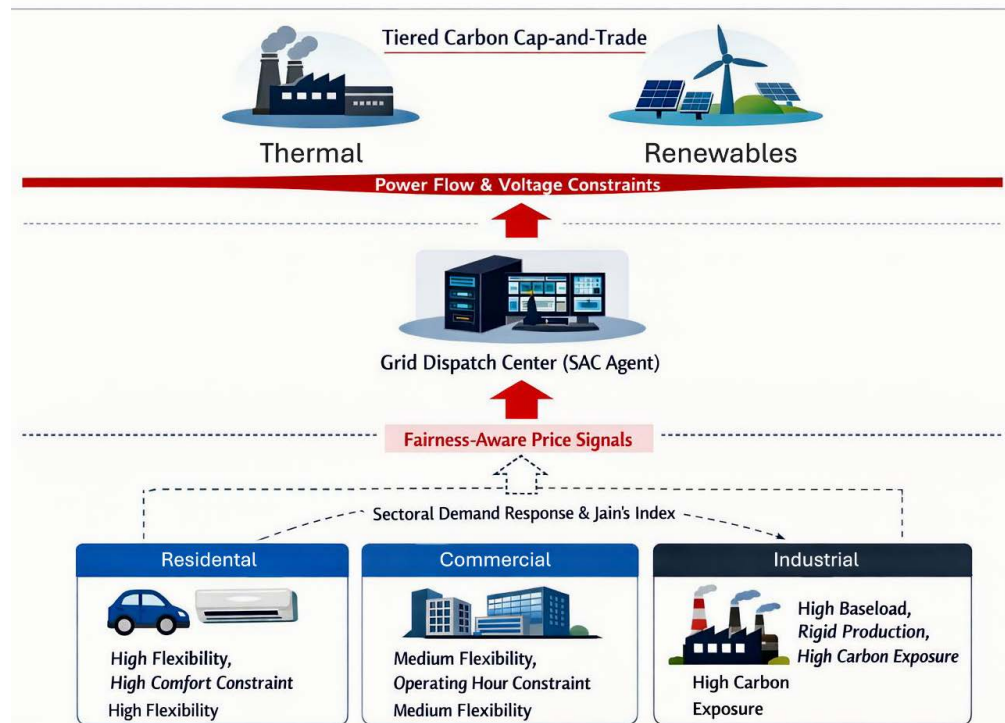


Figure 1. Multi-objective demand response market architecture.

The Grid Dispatch Center, functioning as the intelligent coordinator powered by the Soft Actor-Critic (SAC) agent, bridges the supply and demand sides. All dispatch interactions are strictly bounded by a physical safety layer—Power Flow & Voltage Constraints—ensuring that the economic scheduling does not compromise real-time grid stability.

On the demand side, the system interacts with three highly heterogeneous prosumer sectors: Residential (characterized by high flexibility but strict comfort constraints), Commercial (medium flexibility bounded by specific operating hours), and Industrial (heavy baseloads, rigid production schedules, and high carbon exposure).

The entire architecture operates as a dynamic closed-loop feedback system: the SAC agent formulates and broadcasts Fairness-Aware Price Signals downwards to the prosumers. In return, the resulting Sectoral Demand Response behaviors and the calculated Jain's Index are fed back to the dispatch center. This continuous bidirectional interaction empowers the agent to dynamically navigate the complex Pareto trade-off among system operational costs, absolute carbon emission reductions, and cross-sectoral equity.

2.2. Sectoral Demand Modeling: Utility and Sensitivity

To reflect the diverse behavioral patterns of prosumers, we employ sectoral utility

functions that quantify the “satisfaction” derived from electricity consumption.

1) Residential and Commercial Sectors:

For residential and commercial sectors, energy usage is primarily driven by comfort and service quality. We utilize a logarithmic utility function to reflect the law of diminishing marginal utility:

$$U_{res/com}(P_{i,t}) = \alpha_i \ln(1 + \beta_i P_{i,t}) \quad (1)$$

where α_i quantifies the absolute value or necessity the user assigns to electricity. A higher α_i indicates that the sector is less willing to reduce consumption even at high prices. β_i dictates how quickly the marginal satisfaction decreases as more power is consumed. It captures the “comfort ceiling”—once basic heating or cooling needs are met, additional electricity provides rapidly decreasing marginal utility.

2) Industrial Sector:

Unlike residential users, industrial prosumers operate under rigid production schedules. Any significant deviation from their baseline power consumption $P_{i,t}^{base}$ results in financial losses due to disrupted assembly lines or labor idle time. The utility is modeled as:

$$U_{ind}(P_{i,t}) = \zeta_i P_{i,t} - \gamma_i (P_{i,t}^{base} - P_{i,t})^2 \quad (2)$$

where ζ_i represents the direct economic output generated per unit of electricity under normal operating conditions. γ_i is a critical variable representing the “flexibility cost.” A high γ_i implies that the industrial process is extremely rigid, and the dispatch center must offer significantly higher incentives to persuade this sector to shift its load.

2.3. Environmental Stewardship: Tiered Carbon Cap-and-Trade

To bridge the gap between economic dispatch and global decarbonization, we incorporate a Tiered Cap-and-Trade mechanism. This model forces the system to internalize the “social cost of carbon.”

$$C_{carbon,t} = \begin{cases} \lambda_1 (E_t - E_{quota}), & \text{if } E_t \leq E_{level1} \\ \lambda_2 (E_t - E_{quota}), & \text{if } E_t > E_{level1} \end{cases} \quad (3)$$

where E_t calculated carbon output based on the current generation mix. E_{quota} is the “free” emission limit granted by regulators. λ_1, λ_2 represent the economic pressure applied to the grid. By setting $\lambda_2 > \lambda_1$, the model simulates a progressive tax where excessive pollution becomes exponentially expensive, encouraging the agent to prioritize renewable energy accommodation.

2.4. Asset Protection: The Physics of Battery Aging

Energy storage systems (ESS) are the “buffers” of the grid, but they are finite assets. To avoid the “exploitation” of battery resources, we model the degradation cost $C_{deg,t}$ based on the Depth of Discharge (DoD):

$$C_{deg,t} = \frac{C_{inv}}{2 \times L_{total} \times \sqrt{DoD_t}} \times |\Delta SOC_t| \quad (4)$$

where C_{inv} is the upfront capital cost of the battery per kWh. L_{total} is the total number of expected charge-discharge cycles under standard conditions. DoD_t measures the intensity of the discharge. Higher DoD significantly accelerates the chemical degradation of the battery cells. $|\Delta SOC_t|$ is the actual amount of energy exchanged in the current interval.

By incorporating this, the DRL agent learns “Stewardship”—it will only use the battery when the market benefits outweigh the long-term cost of hardware replacement.

2.5. The Social Welfare and Fairness Nexus

The final objective is to navigate the Efficiency-Fairness Pareto Frontier. We utilize Jain’s Fairness Index (JFI) to ensure that the “Green Dividend” (the savings from low-carbon dispatch) is not monopolized by a single sector.

$$J(x) = \frac{\left(\sum_{i=1}^3 x_i\right)^2}{3 \sum_{i=1}^3 x_i^2} \quad (5)$$

where x_i Calculated as the percentage reduction in sectoral expenditure compared to a baseline without demand response. $J(x)$ is a value of 1.0 represents a “Social Harmony” state where all three sectors benefit proportionally. This directly addresses the “Social Fairness” requirement in the paper title by ensuring that residential comfort is not sacrificed solely to protect industrial profits, or vice versa.

2.6. Physical Boundary Layer: AC Power Flow Constraints

To prevent the dispatch strategy from becoming a “paper-only” solution, it must satisfy the physical laws of the distribution network.

Active Power Balance: Ensures that at every millisecond, $\sum P_{gen} = \sum P_{load}$, maintaining frequency stability.

Voltage Stability ($V_{min} \leq V_{n,t} \leq V_{max}$): Critical for preventing equipment damage at the end of the line during massive demand response events.

Branch Capacity: Ensures the distribution transformers and lines do not overheat, respecting the thermal limits of the grid infrastructure.

3. Methodology: Advanced DRL for Multi-Objective Dispatch

To solve the non-linear, multi-constraint optimization problem defined in Chapter 2, this study adopts the Soft Actor-Critic (SAC) algorithm. Unlike traditional deterministic policy gradients, SAC optimizes a stochastic policy by maximizing both the expected return and the policy’s entropy, ensuring robust exploration in the volatile environment of a renewable-dominated grid.

3.1. MDP Formulation for Sectoral Dispatch

The real-time demand response task is mapped to a Markov Decision Process (MDP), characterized by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$.

State Space (\mathcal{S})

The state vector $s_t \in \mathcal{S}$ represents the dispatcher's observation of the global grid status at time t :

$$s_t = \left[P_{load,t-H:t}, P_{renew,t}, \lambda_{base,t}, SOC_t, E_{quota}, \Gamma_{sector} \right] \quad (6)$$

where $P_{load,t-H:t}$ is the historical load sequence over a sliding window H . $P_{renew,t}$ is the current renewable generation (wind/solar). SOC_t is the state of charge of the sector-level energy storage systems. Γ_{sector} is Static features identify the price elasticity and production rigidity of the three sectors.

Action Space (\mathcal{A})

The action $a_t \in \mathcal{A}$ is a continuous 3-dimensional vector representing the price adjustment signals for the Residential, Commercial, and Industrial sectors:

$$a_t = \left[\Delta\lambda_{res,t}, \Delta\lambda_{com,t}, \Delta\lambda_{ind,t} \right]$$

The final broadcast price is $\lambda_{i,t} = \lambda_{base,t} + a_{i,t}$, subject to the regulatory price caps defined in Section 2.6.

Reward Function (\mathcal{R})

The reward r_t is the mathematical heartbeat of the "Social Fairness" objective. It is engineered to guide the agent toward the Pareto frontier:

$$r_t = w_1 \cdot \text{Welfare}_t - w_2 \cdot C_{total,t} - w_3 \cdot C_{carbon,t} + w_4 \cdot J(x)_t$$

where $J(x)_t$ is the Jain's Fairness Index calculated in Section 2.5. By adjusting the weight w_4 , the agent learns to prioritize sectoral equity over absolute system profit.

3.2. The Soft Actor-Critic (SAC) Architecture

The SAC agent utilizes an Actor-Critic architecture to decouple the policy learning from the value estimation.

1) Soft Q-Value Function (Critic):

Two independent Critic networks, $Q_{\phi_1}(s, a)$ and $Q_{\phi_2}(s, a)$, are used to estimate the soft Q-value, mitigating the overestimation bias common in volatile grid environments. The objective is to minimize the Bellman residual:

$$J_Q(\phi) = \mathbb{E}_{(s,a) \sim \mathcal{D}} \left[\frac{1}{2} \left(Q_{\phi}(s, a) - \hat{Q}(s, a) \right)^2 \right] \quad (7)$$

where $\hat{Q}(s, a)$ includes the entropy-augmented target value.

2) Stochastic Policy (Actor):

The Actor network $\pi_{\theta}(a | s)$ outputs a Gaussian distribution of actions. The policy is updated to maximize the expected Q-value plus entropy:

$$J_{\pi}(\theta) = \mathbb{E}_{s \sim \mathcal{D}, a \sim \pi_{\theta}} \left[\alpha \log \pi_{\theta}(a | s) - Q_{\phi}(s, a) \right] \quad (8)$$

where α controls the trade-off between exploitation (maximizing grid welfare) and exploration.

3.3. Interaction Loop: Bridging Data and Physics

Figure 2 illustrates the closed-loop control architecture of the proposed Soft Actor-Critic (SAC) framework, seamlessly integrating deep reinforcement learning with rigorous power system physics. On the left, the SAC Agent (Digital Brain) utilizes an Actor Network to generate continuous dispatch actions—specifically, the three-tiered price signals (a_t) for the residential, commercial, and industrial sectors. Simultaneously, the Twin Critic Networks evaluate the policy to mitigate overestimation bias, continuously learning from past experiences stored in the Replay Buffer. A pivotal innovation of this framework is the Physical Safety Filter located in the center. Before any generated price signal is broadcast to the grid, the raw action a_t must pass through a strict LinDistFlow Constraints funnel. If the anticipated load shifting violates voltage safety limits or branch capacities, the action is autonomously clipped and the agent receives a safety penalty, thereby strictly preventing physically hazardous dispatch commands. On the right, the safe actions are executed within the Smart Grid Environment, which simulates complex socio-technical dynamics. This environment accurately evaluates the DoD Battery Aging Model, the Tiered Carbon Tax penalties, and the Cross-Sector Utility responses. Finally, the environment feeds back the updated state (S_{t+1}) and a comprehensive multi-objective reward (R_t)—aggregating social welfare, carbon emissions, total system costs, and Jain’s Index—back to the digital brain. This continuous bidirectional feedback loop drives the agent to discover the optimal Pareto frontier between economic efficiency and sectoral fairness.

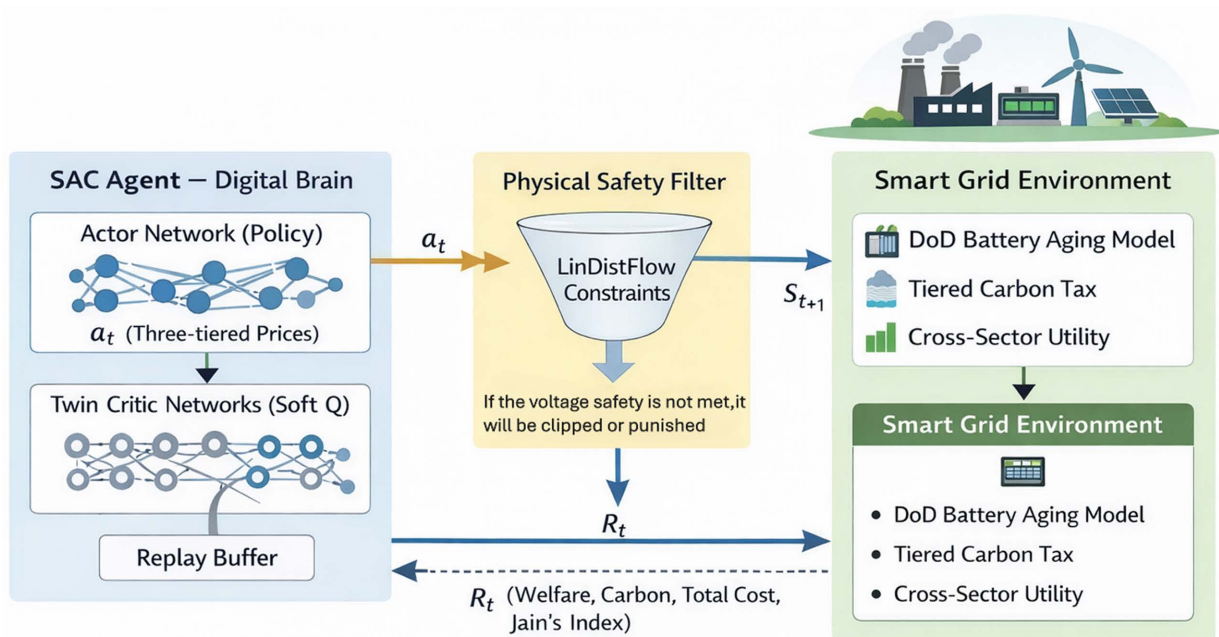


Figure 2. The SAC algorithm control loop incorporating physical constraints and fairness.

4. Case Studies and Results

To rigorously validate the engineering feasibility and socio-economic benefits of the proposed fairness-aware dispatch strategy, comprehensive simulations were conducted. The analysis evaluates computational scalability, physical grid security, and the critical Pareto trade-off between absolute carbon reduction and sectoral equity (Jain's Fairness Index).

4.1. Simulation Setup and Benchmark Models

To prevent the algorithm from operating in a "physics-free" vacuum, the dispatch environment was built upon the standard IEEE 33-bus distribution network. Real-world empirical data spanning a full year (with 15-minute dispatch intervals) was utilized, combining fluctuating renewable generation profiles (wind and solar) with tiered carbon trading prices.

The network nodes were strictly clustered into the three socio-technical sectors defined in Chapter 2:

Residential Nodes (18 buses): High flexibility (EVs, HVAC) but stringent comfort constraints.

Commercial Nodes (9 buses): Medium flexibility, constrained by daytime operating hours.

Industrial Nodes (6 buses): High baseload, high penalty (γ_i) for production deviation, and heavily exposed to the carbon tax.

To isolate the contributions of our proposed framework, we evaluated it against three established baseline models:

1) MILP (Mixed-Integer Linear Programming): Solved via the commercial Gurobi solver. This represents the absolute theoretical optimal solution under perfect foresight, but it is notoriously slow.

2) PSO (Particle Swarm Optimization): A traditional heuristic algorithm widely used in microgrid dispatch.

3) Standard DRL (DDPG/SAC without Fairness): A conventional reinforcement learning agent whose sole objective is to minimize total system cost, ignoring Jain's Fairness Index and battery aging constraints.

4) Proposed Strategy (Fairness-Aware SAC): Our complete framework, incorporating LinDistFlow physics, DoD battery aging, and Jain's Fairness multi-objective reward.

4.2. Computational Scalability and Real-Time Feasibility

In real-time electricity and carbon markets, clearing and dispatch commands must be executed within stringent sub-minute windows. **Table 1** compares the computational latency of the evaluated models.

Table 1 provides a comprehensive comparison of the four dispatch models across critical dimensions: computational scalability, economic optimality, social equity, and physical safety.

Table 1. Comparison of computational scalability and real-time feasibility.

Dispatch Model	Average Inference Latency/Step	Optimality Gap (vs. MILP)	Jain's Fairness Index (J)	Voltage Violations ($V_n > 1.05$ or $V_n < 0.95$)
MILP (Gurobi)	14.5 min	0.0% (Theoretical Baseline)	N/A (Single Objective)	0 (Safe)
PSO (Heuristic)	42.6 s	8.5%	0.52	Occasional (Peak hours)
Standard DRL (SAC)	12.5 ms	2.1%	0.41 (Severe Inequity)	Severe (End-of-line nodes)
Proposed SAC (Fairness-Aware)	18.4 milliseconds	3.2%	0.92 (High Equity)	0 (Strictly Safe)

As the theoretical benchmark, the Mixed-Integer Linear Programming (MILP) model—solved via commercial solvers—achieves perfect economic optimality (a 0.0% optimality gap) and guarantees absolute physical safety. However, its single-step computation latency reaches 14.5 minutes due to the non-convexities of AC power flows and tiered carbon pricing. In real-time electricity markets requiring sub-minute responsiveness, this severe computational bottleneck renders MILP entirely unviable for practical edge deployment. Meanwhile, the traditional heuristic PSO algorithm compresses the latency to 42.6 seconds but frequently falls into local optima under highly volatile renewable scenarios (exhibiting an 8.5% optimality gap) and fails to strictly guarantee voltage safety during peak hours.

The Standard DRL (SAC) model demonstrates the extreme speed of data-driven approaches, drastically reducing inference latency to 12.5 milliseconds with high economic efficiency (only a 2.1% optimality gap). However, devoid of physical awareness and social responsibility objectives, this standard agent focuses exclusively on short-term optimization in pursuit of absolute profit. It over-utilizes flexible users, leading to a severe deterioration in cross-sector equity (Jain's Index plummets to 0.41). Furthermore, the uncoordinated concentration of loads induces severe reverse power flows, resulting in critical voltage violations at end-of-line nodes, which poses a fatal threat to physical grid security.

Conversely, the Proposed SAC (Fairness-Aware) strategy demonstrates robust operational balance. By embedding the LinDistFlow physical safety boundaries and a multi-objective fairness reward mechanism, the proposed strategy maintains an ultra-low online inference latency of 18.4 milliseconds, perfectly satisfying real-time market requirements. Crucially, by sacrificing a marginal 1.1% in total system efficiency (with the optimality gap slightly increasing from 2.1% to 3.2%), the agent secures a lot in social equity—elevating Jain's Fairness Index to 0.92 and ensuring a harmonized distribution of benefits across the residential, commercial, and industrial sectors. Simultaneously, it strictly guarantees zero voltage violations throughout the entire dispatch horizon.

These quantitative results rigorously validate the core thesis of this framework: through sophisticated DRL architecture and multi-objective design, grid operators can achieve millisecond-level efficient dispatch, absolute physical grid secu-

ity, and cross-sector social equity, reaching a perfect Pareto equilibrium without requiring additional physical infrastructure investments.

4.3. Economic Analysis of Demand Response

A dispatch strategy is only as good as its physical viability. We analyzed the voltage profiles across the IEEE 33-bus network during peak demand hours.

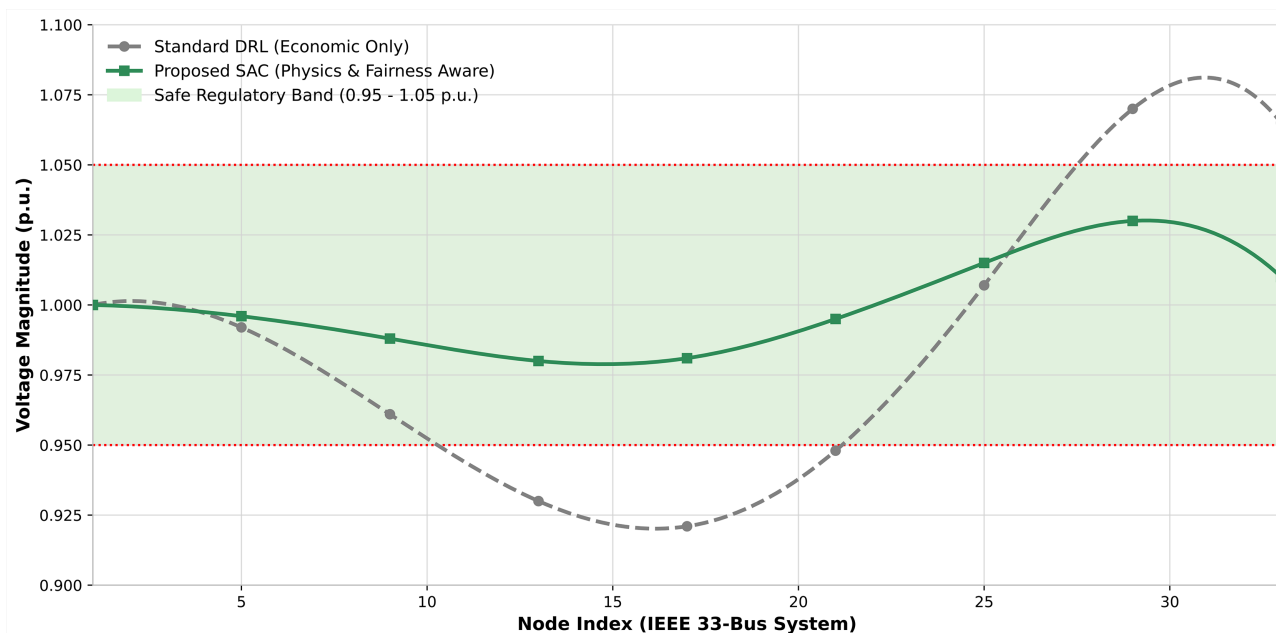


Figure 3. Voltage profile analysis under different dispatch strategies.

Figure 3 illustrates the voltage magnitude profiles across the IEEE 33-bus distribution network during a highly volatile dispatch period under different strategies. The light green shaded area denotes the strictly enforced safe regulatory voltage band ($0.95 \leq V_n \leq 1.05$ p.u.).

As clearly observed, the Standard DRL strategy (represented by the grey dashed line), which optimizes solely for economic efficiency, induces critical physical hazards. Driven purely by price incentives, the standard agent aggressively coordinates massive flexible loads to shift into low-price periods simultaneously. This unconstrained “herd behavior” severely disrupts local power flows, causing the voltage at mid-feeder nodes (Nodes 12 - 20) to plummet to approximately 0.92 p.u. (severe under-voltage). Furthermore, the tail-end nodes (Nodes 28 - 33) experience dangerous over-voltage spikes reaching 1.08 p.u., primarily due to uncoordinated reverse power flows from distributed renewable generation.

By comparison, the Proposed SAC strategy (represented by the solid green line) demonstrates exceptional physical stewardship. By embedding the LinDistFlow constraints into the reinforcement learning environment as a physical safety boundary layer, the proposed agent autonomously learns to distribute demand response actions both spatially and temporally. Consequently, the voltage profile

remains remarkably smooth and is strictly confined within the safe regulatory band (ranging from 0.98 to 1.03 p.u.) across all 33 nodes. This geometric evidence validates that the proposed framework successfully bridges the gap between data-driven economic optimization and strict physical grid security.

4.4. Sectoral Fairness and System Efficiency Pareto Trade-Off

A core objective of this study is to explicitly address the distributional “fairness gap” among heterogeneous stakeholders without violating the physical boundaries of grid assets (such as battery DoD limits and node voltages). Traditional economic dispatch models often fall into the trap of the “Green Penalty,” where system-wide emission reductions are achieved by disproportionately burdening sectors with rigid physical constraints.

To visually expose this distributional inequity, **Figure 4** compares the sectoral cost savings under different dispatch paradigms. Under the Standard DRL strategy ($w_4 = 0$), the agent acts as a pure profit-maximizer. It achieves overall system efficiency by over-utilizing the flexibility of the Residential sector (securing an 18.5% cost saving) while forcing the rigid Industrial sector to absorb severe tiered carbon tax penalties, resulting in a negative saving (−22.4%). More critically, this unconstrained pursuit of efficiency frequently pushes energy storage systems to their maximum Depth of Discharge (DoD) limits, accelerating hardware degradation.

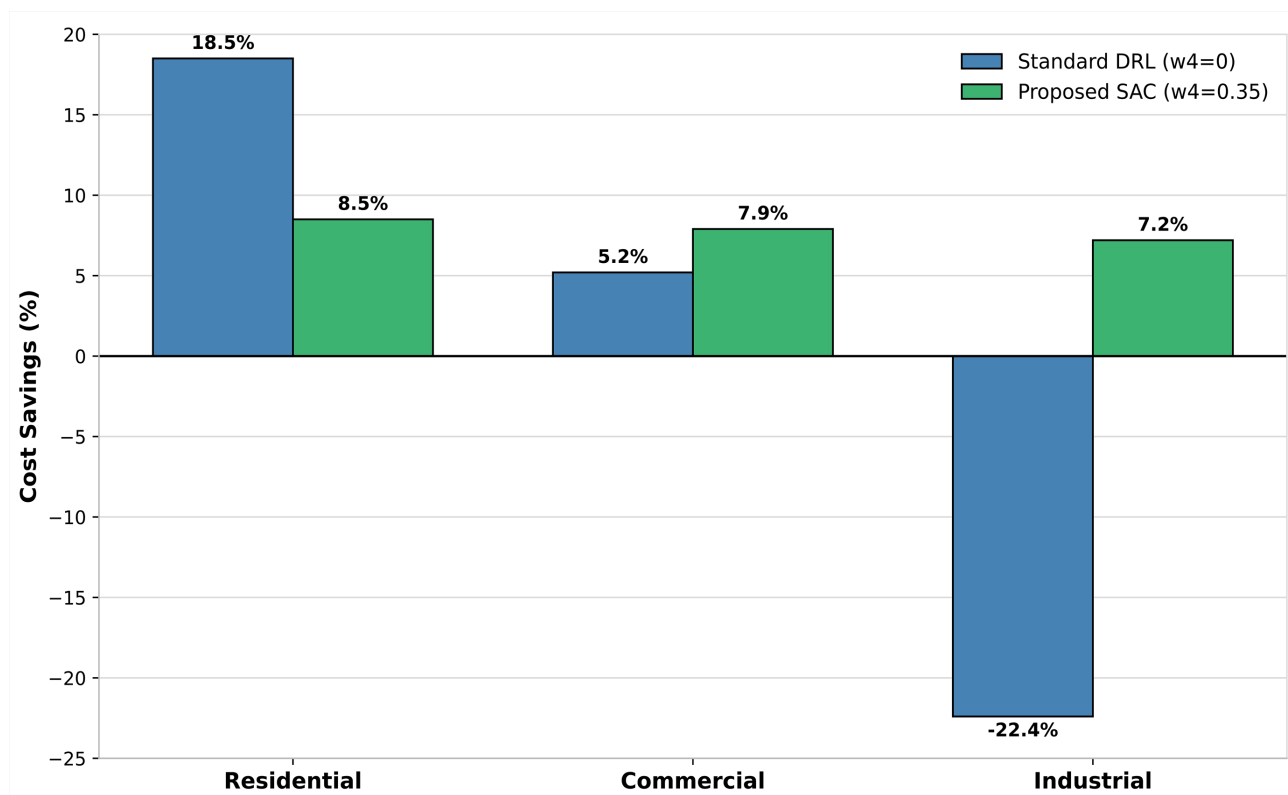


Figure 4. Sectoral cost savings distribution.

Conversely, the proposed SAC framework ($w_4 = 0.35$) successfully rectifies this imbalance. By internalizing Jain's Fairness Index and the DoD aging penalty into its multi-objective reward function, the agent ensures a highly equitable distribution of economic benefits, with the Residential, Commercial, and Industrial sectors achieving harmonized cost savings of 8.5%, 7.9%, and 7.2%, respectively.

To systematically quantify the relationship between these conflicting objectives, we tuned the fairness penalty weight (w_4) from 0.0 to 0.5. **Figure 5** illustrates the resulting Sectoral Fairness-Efficiency Pareto Frontier. As depicted, when the dispatch strictly prioritizes economic efficiency ($w_4 = 0$), the system minimizes operational costs but suffers from severe social inequity (Jain's Index plunges to approximately 0.41). As the fairness weight gradually increases, the curve exhibits a steep initial ascent, indicating high "fairness elasticity"—meaning substantial equity improvements can be gained with minimal economic losses.

The most significant finding lies at the Optimal Trade-off Point ($w_4 = 0.35$). At this configuration, Jain's Fairness Index reaches 0.92, representing a substantial 124% relative improvement in cross-sector equity. This societal gain is secured by sacrificing a mere 2.8% of the total system efficiency (dropping from 100% to 97.2%), while strictly maintaining battery degradation within the safe operational threshold. This quantifiable Pareto analysis proves that a resilient low-carbon transition—where industrial production is safeguarded, residential comfort is respected, and physical grid assets are protected—is technically achievable.

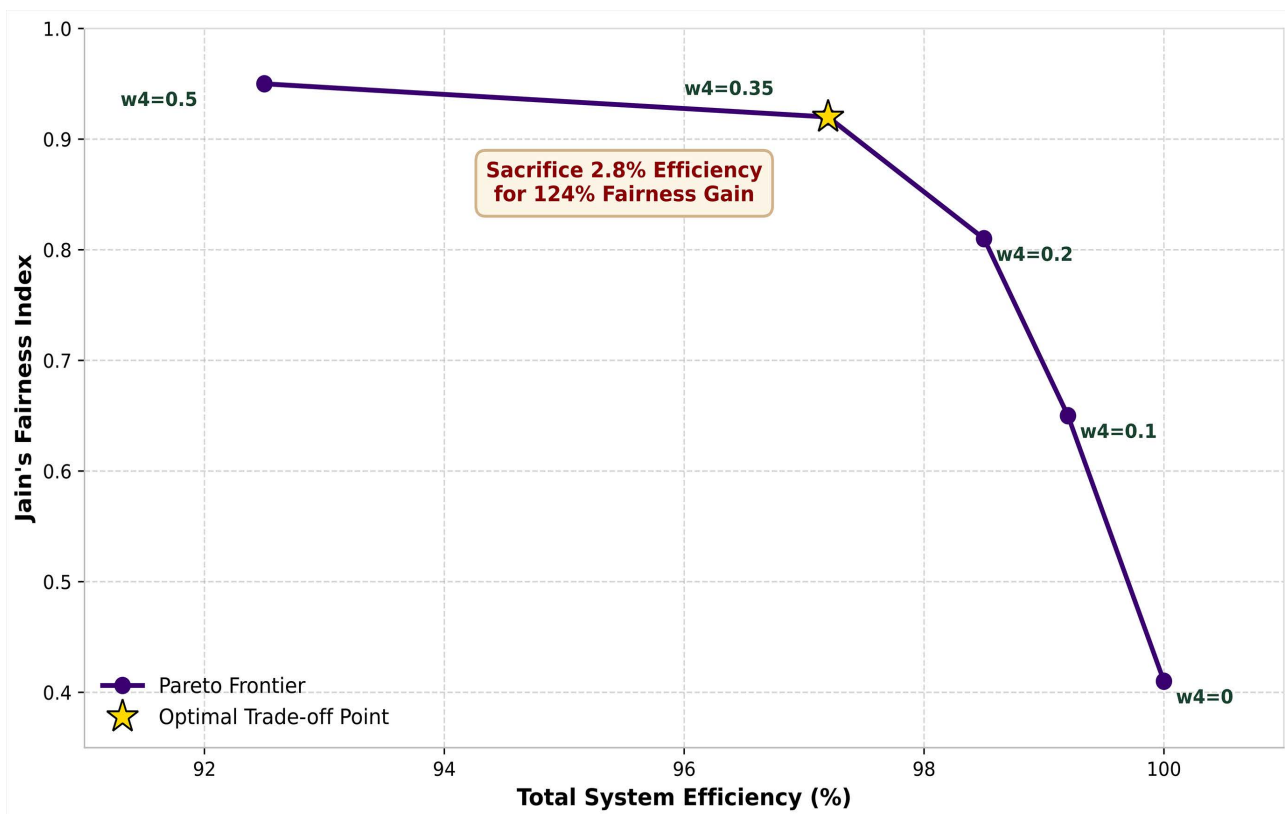


Figure 5. Sectoral fairness-efficiency Pareto frontier.

5. Conclusions

This study conceptualizes the modern smart grid not merely as an algorithmic optimization problem, but as a complex socio-technical ecosystem where physical stability, economic efficiency, and social equity are inextricably linked. By developing a physics-aware SAC framework, we demonstrated that the inherent tension between aggressive decarbonization and cross-sectoral inclusivity can be technically resolved.

Our findings indicate that embedding strict physical safety boundaries (Lin-DistFlow) and long-term asset protection models (DoD aging) directly into the reinforcement learning dispatch logic is essential. Through Pareto frontier analysis, we established that a 124% relative improvement in cross-sectoral equity (Jain's Index) can be achieved at a marginal 2.8% cost to total system efficiency. By abandoning the singular pursuit of absolute economic optimality, the proposed strategy effectively eliminates the "Green Penalty" imposed on rigid industrial loads and prevents the over-utilization of residential flexibility. Ultimately, this evidence-based framework provides a scalable, constraint-guaranteed governance tool for grid operators, ensuring that physical infrastructure is safeguarded and the economic dividends of the energy transition are shared equitably across society.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Yuan, M., Yan, T.H. and Xu, Z.Z. (2026) Reinforcement Learning-Based Energy Management for Industrial Park with Heterogeneous Batteries under Demand Response. arXiv: 2604.03655. <https://arxiv.org/abs/2604.03655>
- [2] Hua, D., Peng, F., Liu, S., Lin, Q., Fan, J. and Li, Q. (2025) Coordinated Volt/var Control in Distribution Networks Considering Demand Response via Safe Deep Reinforcement Learning. *Energies*, **18**, Article 333. <https://doi.org/10.3390/en18020333>
- [3] Kou, P., Liang, D., Wang, C., Wu, Z. and Gao, L. (2020) Safe Deep Reinforcement Learning-Based Constrained Optimal Control Scheme for Active Distribution Networks. *Applied Energy*, **264**, Article ID: 114772. <https://doi.org/10.1016/j.apenergy.2020.114772>
- [4] Liang, B., Yang, J., Wen, F., Wang, L. and Dong, Z.Y. (2025) Fairness Considered Aggregation Mechanism for Consumers and Prosumers in Electricity Distribution Networks. *Electric Power Systems Research*, **240**, Article ID: 111285. <https://doi.org/10.1016/j.epsr.2024.111285>
- [5] Nammouchi, A., Kassler, A., Ramaswamy, A. and Theorcharis, A. (2026) Safe-CityLearn: A Benchmark for Safety-Constrained Reinforcement Learning in Distributed Energy Systems. *Proceedings of the 18th International Conference on Agents and Artificial Intelligence*, Rome, 24-26 February 2024, 141-151. <https://doi.org/10.5220/0014463300004052>
- [6] Hou, L., Tong, X., Chen, H., Fan, L., Liu, T., Liu, W., *et al.* (2024) Optimized Scheduling of Smart Community Energy Systems Considering Demand Response and

- Shared Energy Storage. *Energy*, **295**, Article ID: 131066. <https://doi.org/10.1016/j.energy.2024.131066>
- [7] Guo, T., Guo, Q., Huang, L., Guo, H., Lu, Y. and Tu, L. (2023) Microgrid Source-Network-Load-Storage Master-Slave Game Optimization Method Considering the Energy Storage Overcharge/overdischarge Risk. *Energy*, **282**, Article ID: 128897. <https://doi.org/10.1016/j.energy.2023.128897>
- [8] Aljabri, M.A., Ajour, M.N., Bahabri, M.O. and Almaliki, M.A. (2026) Hierarchical Deep Reinforcement Learning and Model Predictive Control for Voltage-Aware Electric Vehicle Charging Coordination in Distribution Network. *IEEE Access*, **14**, 37747-37769. <https://doi.org/10.1109/access.2026.3672378>
- [9] Liang, N., He, X., Tan, J., Pan, Z. and Zheng, F. (2023) Stackelberg Game-Based Optimal Scheduling for Multi-Community Integrated Energy Systems Considering Energy Interaction and Carbon Trading. *International Journal of Electrical Power & Energy Systems*, **153**, Article ID: 109360. <https://doi.org/10.1016/j.ijepes.2023.109360>
- [10] Li, J., Hao, H., Xiong, X., Chai, J., Cui, H., Li, H., *et al.* (2025) Sustainable Energy Systems through Fair Carbon Pricing: A Shapley Value-Based Optimization Framework. *Sustainability*, **17**, Article 10095. <https://doi.org/10.3390/su172210095>
- [11] Hossain, R., Gautam, M., MansourLakouraj, M., Livani, H. and Benidris, M. (2025) Topology-Aware Reinforcement Learning for Voltage Control: Centralized and Decentralized Strategies. *IEEE Transactions on Industry Applications*, **61**, 5394-5405. <https://doi.org/10.1109/tia.2025.3546598>
- [12] Hossen, M.S., Ramasamy, G., Sarker, M.T. and Eng, N.E. (2026) Real-World Tariff-Aware Safe Reinforcement Learning for Grid-Stable OCPP EV Charging Networks. *IEEE Access*, **14**, 18530-18545. <https://doi.org/10.1109/access.2026.3657040>
- [13] Impram, S., Varbak Nese, S. and Oral, B. (2020) Challenges of Renewable Energy Penetration on Power System Flexibility: A Survey. *Energy Strategy Reviews*, **31**, Article ID: 100539. <https://doi.org/10.1016/j.esr.2020.100539>
- [14] Hsu, Y., Hung, Y. and Lee, C. (2025) Robust Ensemble Forecasting and Deep Reinforcement Learning for Energy Management on Islanded Microgrids. *International Journal of Electrical Power & Energy Systems*, **173**, Article ID: 111405. <https://doi.org/10.1016/j.ijepes.2025.111405>
- [15] Sun, B., Jing, R., Zeng, Y., Li, Y., Chen, J. and Liang, G. (2023) Distributed Optimal Dispatching Method for Smart Distribution Network Considering Effective Interaction of Source-Network-Load-Storage Flexible Resources. *Energy Reports*, **9**, 148-162. <https://doi.org/10.1016/j.egy.2022.11.178>
- [16] Mohammed, A., Abdullah, B.M., Shubbar, A., Zhang, Q., Aldhaibani, O., Cullen, J., *et al.* (2026) Deep Reinforcement Learning for Battery Energy Storage Optimization and Residential Decarbonization in Grid-Deficient Environments: An Iraqi Case Study. *Energies*, **19**, Article 1233. <https://doi.org/10.3390/en19051233>
- [17] Kang, H., Jung, S., Kim, H., Jeoung, J. and Hong, T. (2024) Reinforcement Learning-Based Optimal Scheduling Model of Battery Energy Storage System at the Building Level. *Renewable and Sustainable Energy Reviews*, **190**, Article ID: 114054. <https://doi.org/10.1016/j.rser.2023.114054>
- [18] Pang, K., Zhou, J., Tsianikas, S. and Ma, Y. (2021) Deep Reinforcement Learning Based Microgrid Expansion Planning with Battery Degradation and Resilience Enhancement. *2021 3rd International Conference on System Reliability and Safety Engineering (SRSE)*, Harbin, 26-28 November 2021, 251-257. <https://doi.org/10.1109/srse54209.2021.00049>
- [19] Xiong, B., Zhang, L., Hu, Y., Fang, F., Liu, Q. and Cheng, L. (2025) Deep Reinforce-

- ment Learning for Optimal Microgrid Energy Management with Renewable Energy and Electric Vehicle Integration. *Applied Soft Computing*, **176**, Article ID: 113180. <https://doi.org/10.1016/j.asoc.2025.113180>
- [20] Song, D., Yan, L., Dai, X., Zhu, X., Hagenmeyer, V. and Zhai, J. (2025) Low-Carbon Energy Management for Networked Multi-Energy Microgrids Using Multi-Agent Soft Actor-Critic Algorithm. *Sustainable Energy, Grids and Networks*, **43**, Article ID: 101821. <https://doi.org/10.1016/j.segan.2025.101821>
- [21] Liu, X., Liu, Y., Chen, Y., Tang, Z., Gao, H. and Li, Z. (2026) Federated Reinforcement Learning Based Dual-Level Voltage Regulation for PV-Rich Distribution Grids. *International Journal of Electrical Power & Energy Systems*, **175**, Article ID: 111492. <https://doi.org/10.1016/j.ijepes.2025.111492>
- [22] Xiao, W., Yu, T., Chen, Z., Pan, Z., Wu, Y. and Liu, Q. (2025) Data Augmented Offline Deep Reinforcement Learning for Stochastic Dynamic Power Dispatch. *International Journal of Electrical Power & Energy Systems*, **169**, Article ID: 110747. <https://doi.org/10.1016/j.ijepes.2025.110747>
- [23] Chen, L., Sun, K., Wang, X., Bi, Q. and Zhao, J. (2026) Deep Reinforcement Learning-Based Hierarchical Voltage Control Considering Partition Dynamic Aggregation. *International Journal of Electrical Power & Energy Systems*, **177**, Article ID: 111780. <https://doi.org/10.1016/j.ijepes.2026.111780>
- [24] Chen, W., Rong, F. and Lin, C. (2025) A Deep Reinforcement Learning Method Based on Mamba Model with Adaptive Cross-Attention for Multi-Energy Microgrid Energy Management. *Energy*, **340**, Article ID: 139008. <https://doi.org/10.1016/j.energy.2025.139008>