

Reinforcement Learning for Vehicle-to-Grid Voltage Regulation: Single-Hub to Multi-Hub Coordination with Battery-Aware Constraints

Jingbo Wang, *Graduate Student Member, IEEE*, Roshni Anna Jacob, *Member, IEEE*,
Harshal D. Kaushik, *Member, IEEE*, Jie Zhang, *Senior Member, IEEE*
The University of Texas at Dallas, Richardson, Texas 75080, USA
Email: jiezhang@utdallas.edu

Abstract—This paper presents a Vehicle-to-Grid (V2G) coordination framework using reinforcement learning (RL). An intelligent control strategy based on the soft actor-critic algorithm is developed for voltage regulation through single and multi-hub charging systems while respecting realistic fleet constraints. A two-phase training approach integrates stability-focused learning with battery-aware deployment to ensure practical feasibility. Simulation studies on the IEEE 34-bus system validate the framework against a standard Volt-Var/Volt-Watt droop controller. Results indicate that the RL agent achieves performance comparable to the baseline control strategy in nominal scenarios. Under aggressive overloading, it provides robust voltage recovery (within 10% of the baseline) while prioritizing fleet availability and state-of-charge preservation, demonstrating the viability of constraint-aware learning for critical grid services.

Index Terms—Vehicle-to-grid, electric vehicle fleet, voltage regulation, distribution networks, reinforcement learning, soft actor-critic, multi-hub coordination.

I. INTRODUCTION

The rapid proliferation of electric vehicles (EVs) is transforming distribution grids, introducing voltage challenges while opening possibilities for Vehicle-to-Grid (V2G) services. Bidirectional V2G systems enable EVs to function as distributed energy resources (DERs), modulating active and reactive power to support voltage regulation [1]. In this context, we focus on EV “fleets” (e.g., logistics or ride-sharing entities) where a central operator controls charging and discharging actions across either a single hub or a multi-hub system. However, effectively harnessing this flexibility requires sophisticated control frameworks that integrate network constraints with realistic battery dynamics.

Traditional voltage regulation relies on legacy devices with limited actuation speeds and discrete settings, rendering them inadequate for rapid fluctuations. In contrast, inverter-based DERs are increasingly gaining attention in this regard, as they can be controlled to offer continuous response with sub-second response times. In parallel, reinforcement learning (RL) has emerged as a promising solution for developing intelligent control strategies. RL has been applied to various critical grid operations, from real-time outage management [2], [3] to distribution grid voltage control [4], [5]. While these studies demonstrate RL’s capabilities, existing works on voltage

regulation often model controllable resources with static capacity limits, neglecting temporal or state-dependent flexibility constraints inherent in V2G systems [6], [7].

Despite advancements in RL-based voltage control, several critical gaps remain in addressing realistic V2G system constraints. Existing studies predominantly focus on single aggregators or homogeneous fleets, leaving the coordination of spatially distributed charging hubs (within a feeder) operating under unified control policies largely underexplored. Moreover, most distribution network control models assume static battery capacity bounds and neglect key factors such as state of charge (SOC) or state of health (SOH) dynamics, as well as continuously derated power envelopes that arise with real-world EV deployment. While existing studies focus on either single-hub or multi-hub scenarios in isolation, the scalability challenges of expanding from single-hub to multi-hub V2G systems are not adequately addressed. The aggregated behavior of EV fleets is influenced by heterogeneous factors SOC, SOH, availability patterns, and nonlinear battery constraints, which must be integrated into control frameworks for practical deployment [8], [9].

Building on prior work in EV infrastructure planning that addressed optimal charging station deployment considering hosting capacity and voltage constraints [10], [11], this paper shifts focus to operation, aiming to utilize EV flexibility to improve voltage regulation. To address these gaps, this paper proposes a V2G control architecture that enables both single-hub and multi-hub coordination, employing RL-based intelligent control to dispatch active and reactive power across multiple geographically distributed hubs. A fleet-aware power mapping module translates hub-level power signals into battery-level actions by incorporating power conversion, inverter efficiency, and SOC/SOH-dependent bounds.

The remainder of this paper is organized as follows: Section II presents the methodology, including the system overview, distribution network hub, EV fleet models, and the RL framework. Section III describes the case study setup and RL agent training environment. Section IV presents the simulation results, providing a comparative analysis against industry-standard Volt-Var/Volt-Watt local control strategies to validate the proposed framework’s efficacy in voltage regulation

while strictly adhering to fleet constraints. Finally, Section V concludes the paper with key findings and future research directions.

II. METHODOLOGY

This section presents the methodological framework for voltage regulation through coordinated V2G control using RL, while integrating power flow simulation and realistic EV fleet modeling with the learning-based control.

A. System Overview and Hub Model

The system under consideration comprises a radial distribution feeder integrated with V2G hubs at select buses. Each hub aggregates EVs that interface bidirectionally with the grid through smart inverters. The active and reactive power setpoints (P, Q) at each hub can be controlled by an agent. In single-hub scenarios, one V2G hub is controlled for regulation, while multi-hub scenarios coordinate multiple hubs simultaneously. The RL framework includes: (i) OpenDSS-based power flow environment, (ii) RL agent for optimal hub-level control decisions, and (iii) EV fleet layer translating control outputs into physically realizable power injections subject to SOC, SOH, and inverter efficiency constraints.

The distribution network is simulated using OpenDSS where each V2G hub (h) is modeled as a controllable three-phase generator. The apparent power at the hub is represented as $S^{(h)} = P^{(h)} + jQ^{(h)}$, subject to active and reactive power limits $|P^{(h)}| \leq P_{\max}^{(h)}$ and $|Q^{(h)}| \leq Q_{\max}^{(h)}$.

B. EV Fleet Model and Operational Constraints

Each V2G hub aggregates a fleet of EVs whose individual states collectively determine available charging and discharging capability. The fleet model integrates electrochemical limits, inverter efficiency, and participation constraints to ensure physically consistent operation.

1) *Battery Power Constraints:* Each EV (e) is characterized by its battery capacity (C_e) , state of charge (SOC_e) , and state of health (SOH_e) . Each EV's instantaneous power capability P_e^{bat} is estimated based on its predicted operating conditions:

$$P_e^{\text{bat}} = \min(P_{\text{voltage},e}, P_{\text{current},e}) \quad (1)$$

where $P_{\text{voltage},e}$ and $P_{\text{current},e}$ represent voltage-limited and current-limited power based on SOC and SOH. The power constraints are enforced through C-rate limitations that vary with SOC and SOH, and inverter constraints, ensuring realistic battery behavior during grid service operations.

2) *SOC and SOH Evolution:* Battery states evolve based on current throughput and degradation factors:

$$\begin{aligned} SOC_e(t + \Delta t) &= SOC_e(t) + \frac{I_e^{\text{bat}} \Delta t}{C_e SOH_e}, \\ SOH_e(t + \Delta t) &= SOH_e(t) - \Delta SOH_{\text{cycle}} - \Delta SOH_{\text{cal}}, \end{aligned} \quad (2)$$

where $I_e^{\text{bat}} = P_e^{\text{bat}}/V_e$ is the battery current (positive for charging, negative for discharging), V_e is the battery voltage in V, and degradation depends on throughput energy, temperature, and duty cycle. $\Delta SOH_{\text{cycle}}$ represents cycle-based degradation

from energy throughput and ΔSOH_{cal} represents calendar aging degradation from time and temperature effects.

3) *Fleet-Level Power Allocation:* At the fleet level, the grid power injection required for voltage regulation is translated to behind-the-meter fleet demand using apparent power and inverter efficiency considerations. The total apparent power requested for regulation at the hub is:

$$S_{\text{req}}^{(h)} = \sqrt{P_{\text{req}}^{(h)2} + Q_{\text{req}}^{(h)2}} \quad (3)$$

where P_{req} and Q_{req} denote the active and reactive power injection required at the hub for voltage regulation, respectively. The battery power required to meet this requested demand is:

$$P_{\text{fleet}}^{(h)} = \frac{S_{\text{req}}^{(h)}}{\eta_{\text{inv}}} \quad (4)$$

where η_{inv} is the inverter efficiency. When the requested power exceeds fleet capability $P_{\text{avail}}^{(h)}$, a proportional scaling ensures feasibility:

$$\rho^{(h)} = \min \left(1, \frac{P_{\text{avail}}^{(h)}}{P_{\text{fleet}}^{(h)}} \right), \quad (5)$$

and the actual power delivered to the grid is:

$$P_{\text{sup}}^{(h)} = \rho^{(h)} P_{\text{fleet}}^{(h)}, \quad Q_{\text{sup}}^{(h)} = \rho^{(h)} Q_{\text{fleet}}^{(h)}. \quad (6)$$

This hierarchical allocation mechanism ensures that fleet-level power injections remain physically consistent with individual EV capabilities, inverter limits, and SOH-dependent availability.

C. Reinforcement Learning Framework

The voltage regulation task is formulated as a Markov Decision Process (MDP) defined by $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r)$. The OpenDSS environment serves as the simulation environment, encapsulating network equations, voltage constraints, and dynamic interactions between load conditions and V2G power injections.

1) *State and Action Spaces:* The state space $\mathcal{S}$ consists of: (i) bus voltage magnitudes V_i^{pu} for all monitored buses in per unit (p.u.). The action space $\mathcal{A}$ contains continuous power scaling factors for hub injections, where the agent's action vector is $\mathbf{a} = [a_P^{(1)}, a_Q^{(1)}, \dots, a_P^{(h)}, a_Q^{(h)}]$ for h hubs, which are normalized control factors bounded in $[-1, 1]$.

These actions are converted to real and reactive power setpoints:

$$P_{\text{req}}^{(h)} = a_P^{(h)} P_{\max}, \quad Q_{\text{req}}^{(h)} = a_Q^{(h)} Q_{\max}, \quad (7)$$

2) *Reward Function:* The reward function guides the RL model to regulate voltage in the network through the V2G control:

$$r_t = R_{\text{vb}} - R_{\text{vp}} \quad (8)$$

where:

$$R_{\text{vb}} = \begin{cases} 10.0 & \text{if all voltages in range} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$R_{\text{vp}} = \sum_i \begin{cases} (V_{\min} - V_i) \times 100 & \text{if } V_i < V_{\min} \\ (V_i - V_{\max}) \times 100 & \text{if } V_i > V_{\max} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

The voltage limits are the traditional thresholds considered in power systems, i.e., $V_{\min} = 0.95$ p.u. and $V_{\max} = 1.05$ p.u.

3) *SAC Algorithm*: The agent employs the Soft Actor-Critic (SAC) algorithm, an entropy-regularized actor-critic method well-suited for continuous control problems. SAC maximizes the expected cumulative reward augmented with an entropy term to promote exploration:

$$J(\pi) = \mathbb{E}_{(s_t, a_t) \sim \mathcal{D}} [Q_\theta(s_t, a_t) - \alpha \log \pi_\phi(a_t|s_t)], \quad (11)$$

Here, $\pi_\phi(a_t|s_t)$ denotes the stochastic actor (policy) network parameterized by ϕ , which outputs a probability distribution over actions given state s_t , and $Q_\theta(s_t, a_t)$ represents the critic network parameterized by θ , which estimates the expected return for each state-action pair. The temperature coefficient α balances reward maximization and policy stochasticity, controlling the trade-off between exploitation and exploration. The replay buffer $\mathcal{D}$ stores past transitions (s_t, a_t, r_t, s_{t+1}) used for off-policy learning.

D. Training and Deployment Workflow

The methodology employs a two-phase workflow shown in Fig. 1. Phase 1 trains the agent in an idealized environment with fixed hub power limits and no explicit fleet constraints, using time-varying load conditions through load multipliers $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ applied to base loads: $P_i^{\text{load}}(\lambda) = \lambda P_i^{\text{base}}$, $Q_i^{\text{load}}(\lambda) = \lambda Q_i^{\text{base}}$. Phase 2 evaluates the trained policy with detailed fleet model enabled, where hub-level scaling ratio $\rho^{(h)}$ adjusts agent outputs based on real-time fleet availability while SOC and SOH states evolve dynamically. This two-phase design implements policy learning and physical EV constraints enforcement in different modules, ensuring operational feasibility and training stability.

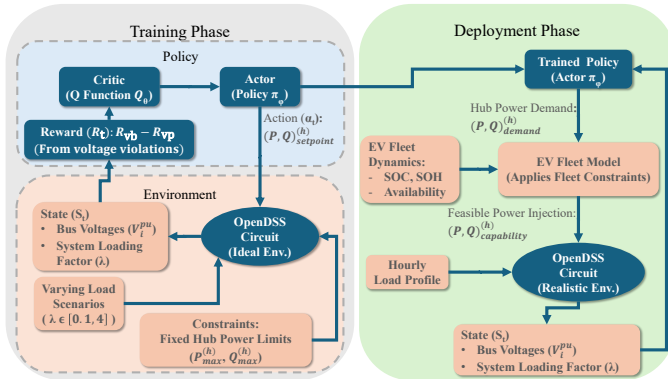


Fig. 1: Learning architecture and training-deployment workflow for V2G voltage regulation control.

III. CASE STUDY

A. Simulation Environment

The proposed framework is validated on the IEEE 34-bus radial distribution feeder, modeled in OpenDSS and interfaced with the Gymnasium RL framework via OpenDSSDirect [?]. The feeder operates at a nominal voltage of 24.9 kV, with base loading of 1.8 MW and 1 MVar.

In the training phase, each episode comprises 100 discrete control steps with randomly sampled load multipliers $\lambda \in [0.1, 4.0]$ to ensure exposure to diverse operating conditions. For deployment evaluation, both mild and aggressive overloading scenarios are considered, resulting in minimum voltages of 0.907 p.u. and 0.807 p.u., respectively, in the absence of regulation. The daily evaluation horizon uses 1-hour discrete steps, with V2G services active from 6:00 to 23:00, as shown in Fig. 2.

For single-hub evaluation, one V2G hub is connected at bus 890. For multi-hub coordination, five V2G hubs are connected at buses 890, 844, 832, 830, and 860. Hubs have rated power of 500 kW (active) and 400 kVar (reactive) each. Each EV has a 75 kWh battery, initial SOC range of 0.2 to 0.9, SOH of 0.95, charging/discharging efficiency of 0.96, and C-rate limit of 0.5C. The single-hub scenario represents a single fleet hub with realistic EV availability patterns, while the multi-hub scenario represents coordination across multiple independent fleet hubs where the aggregator coordinates power dispatch across all hubs.

For the single-hub case study, time-varying EV availability schedules are incorporated to model realistic fleet participation patterns, where the number of available EVs varies hourly based on delivery schedules (45-85% availability). For the multi-hub case study, the evaluation focuses on coordination capabilities by assuming sufficient EV capacity at each hub, allowing the analysis to isolate the benefits of multi-hub coordination from availability constraints.

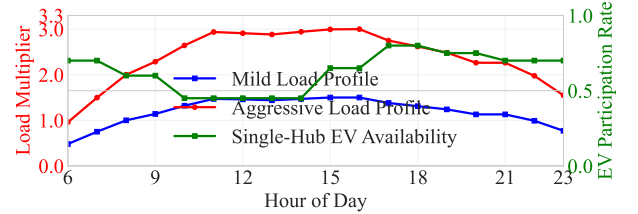


Fig. 2: **Evaluation Profiles.** The left axis displays the time-varying load multipliers for both the **Mild** (peak $\lambda = 1.5$) and **Aggressive** (peak $\lambda = 3.0$) scenarios. The right axis indicates the corresponding EV fleet participation rate during the active V2G window (06:00–23:00).

B. Control Algorithms

1) *RL Agent*: The control agent utilizes the Soft Actor-Critic (SAC) algorithm. The actor and critic networks each consist of two hidden layers with 256 ReLU units. Key hyperparameters include a learning rate of 3×10^{-4} , batch size of 256, discount factor $\gamma = 0.99$, and replay buffer size of 10^6 .

2) *Baseline Strategy*: To validate performance, we benchmark against a decentralized Droop control strategy. Each hub implements autonomous piecewise-linear Volt-Var ($Q(v)$) and Volt-Watt ($P(v)$) curves with a deadband of ± 0.02 p.u. and saturation at $[0.90, 1.10]$ p.u. This represents a state-of-the-

TABLE I: Single-hub voltage regulation summary. Statistics are computed from hourly mean bus voltages across the feeder; violation hours count hours with mean voltage < 0.95 p.u.

Scenario	Mild loading				Aggressive loading			
	Mean	Min	Max	Viol.	Mean	Min	Max	Viol.
Baseline	0.963	0.907	1.048	13	0.883	0.807	1.017	17
RL (no EV)	0.989	0.942	1.048	6	0.896	0.812	1.017	17
Droop (no EV)	0.987	0.943	1.048	6	0.898	0.821	1.017	17
RL (EV-constr.)	0.969	0.910	1.048	12	0.888	0.810	1.017	17
Droop (EV-constr.)	0.972	0.908	1.048	11	0.889	0.807	1.017	17

art local control baseline where inverters act independently without inter-hub communication.

IV. RESULTS

A. Single-Hub Voltage Regulation Performance

This section evaluates the voltage regulation capability of a single V2G hub under mild and aggressive loading conditions, with the objective of distinguishing controller capability from realistic deployment limits. To this end, we first report results without EV constraints to isolate how effectively RL and droop control can support voltage when sufficient energy is available, and then enforce EV constraints to quantify how availability and SOC limitations reduce achievable performance. Mild loading is presented first as a representative undervoltage case, followed by aggressive loading to illustrate how single-hub performance degrades under extreme stress.

Table I summarizes the single-hub voltage regulation results across both loading profiles. Under mild loading and without EV constraints, both RL and droop substantially improve voltage conditions relative to the baseline, reducing violation hours from 13 to 6 and achieving similar minimum voltages. When EV constraints are enforced, the achievable gains are significantly reduced for both controllers, with violation hours remaining close to the baseline level. In this constrained setting, RL and droop exhibit comparable performance, indicating that fleet availability and SOC limits, rather than the specific control strategy, dominate single-hub voltage regulation outcomes. These results suggest that, for a single hub, fleet availability rather than inverter rating becomes the primary performance bottleneck once realistic constraints are applied. Under aggressive loading, voltage violations persist across all single-hub configurations, regardless of control strategy or constraint setting. While both RL and droop provide marginal improvements in mean and minimum voltage relative to the baseline in the unconstrained case, enforcing EV constraints eliminates most of these gains, with violation hours remaining unchanged at 17.

Figure 3 provides time-series context for these results. The voltage improvements concentrate during periods when EV availability and SOC headroom are highest, and diminish as participation decreases or batteries approach lower SOC levels. In the aggressive case, deeper SOC drawdown and tighter availability further constrain support, consistent with the limited improvements observed in Table I. Overall, these results

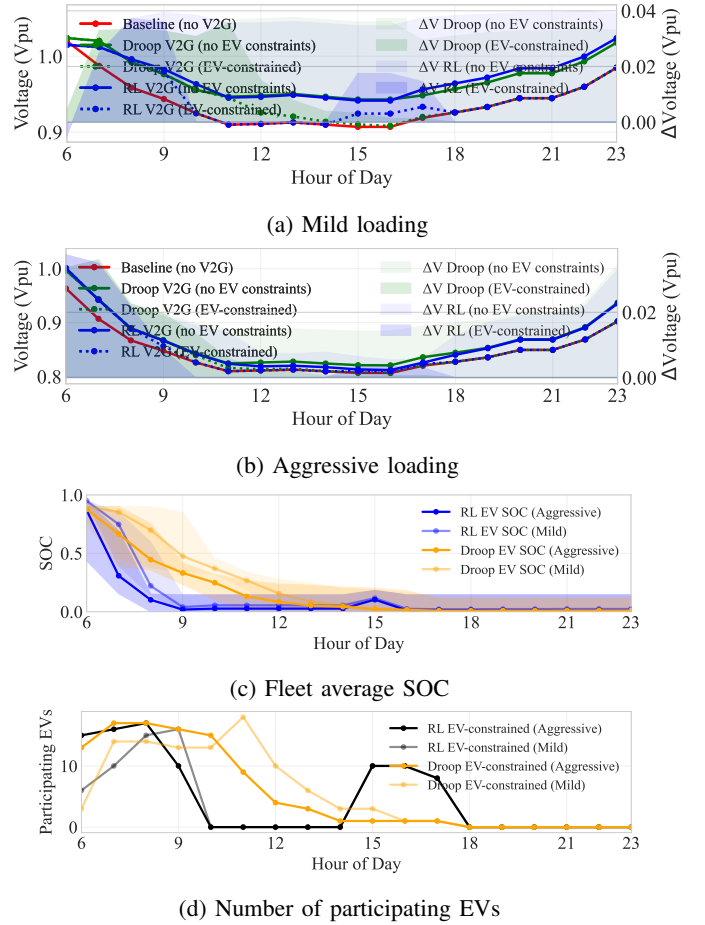


Fig. 3: **Single-Hub Voltage Regulation Performance.** (a)–(b) Mean feeder voltage profiles and improvements under mild and aggressive loading. (c) Fleet-average SOC. (d) Participating EV count. Voltages are averaged across all buses.

demonstrate that while single-hub V2G control can improve voltage margins under moderate conditions, its effectiveness is fundamentally constrained by topology and availability, motivating the need for coordinated multi-hub support.

B. Multi-Hub Coordination

We evaluate coordinated control across five hubs under mild and aggressive loading to benchmark against droop and to assess generalization and stress response. The mild case reflects typical undervoltage operation, while the aggressive case stresses voltage recovery when feeder-wide support is required.

Table II summarizes the coordinated multi-hub results, with time-series behavior illustrated in Fig. 4a and Fig. 4b.

Under mild loading, both coordinated RL and droop control eliminate voltage violations and achieve nearly identical mean and minimum voltages, indicating that moderate voltage deviations can be effectively handled by either strategy. The RL policy maintains voltages close to nominal without introducing over-voltage risks, suggesting that it captures the underlying

TABLE II: Multi-hub voltage regulation summary. Statistics are computed from hourly mean bus voltages; violation hours count hours with mean voltage < 0.95 p.u.

Scenario	Mild loading				Aggressive loading			
	Mean	Min	Max	Viol.	Mean	Min	Max	Viol.
Baseline	0.963	0.907	1.048	13	0.883	0.807	1.017	17
RL (coord.)	1.025	1.007	1.048	0	0.949	0.903	1.017	15
Droop (coord.)	1.024	1.004	1.048	0	0.998	0.940	1.035	2

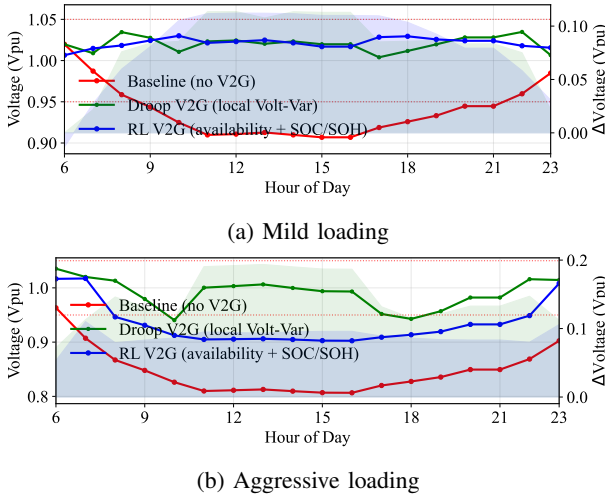


Fig. 4: **Multi-Hub Voltage Regulation.** (a) Mild loading and (b) aggressive loading scenarios under coordinated V2G control. Subplots show mean voltages averaged across all network buses, illustrating feeder-wide voltage response relative to the baseline.

voltage sensitivity of the network rather than overfitting to a specific loading trajectory.

Under aggressive loading, the uncontrolled baseline exhibits sustained feeder-wide voltage sags, while both coordinated controllers substantially improve the voltage profile, confirming the importance of spatial coordination across multiple hubs when single-point support is insufficient. In this regime, the droop controller attains higher mean and minimum voltages and reduces violation hours more effectively by aggressively driving inverter outputs toward their limits based on local voltage measurements. The RL controller does not match droop performance in this stress case, but still delivers a consistent and nontrivial voltage uplift across the feeder through coordinated multi-hub dispatch. Overall, while rule-based droop remains highly effective for saturation-driven voltage correction under extreme conditions, the coordinated RL framework demonstrates a flexible and extensible approach for learning feeder-wide coordination, providing a promising foundation for future integration of fleet-aware constraints and additional system-level objectives.

V. CONCLUSION

This paper presents a V2G coordination framework for voltage regulation using reinforcement learning, demonstrating both single-hub and multi-hub coordination capabilities. The framework introduces a two-phase operational approach with policy learning based on charger hub integrated distribution feeder model and deployment module that enforces the realistic fleet constraints, facilitating training stability while ensuring practical feasibility.

The multi-hub results demonstrate that coordinated RL can provide meaningful feeder-wide support, but a local droop baseline can still outperform it under aggressive stress. This motivates future work on constraint-aware objectives, including battery-degradation-aware optimization, and extensions to larger feeders, multi-agent coordination, and the integration of vehicle travel and logistics constraints.

ACKNOWLEDGMENT

This work was partially supported by the Department of the Navy, Office of Naval Research under ONR award number N00014-21-1-2530 and National Science Foundation under Award 2229417.

REFERENCES

- [1] S. Alamgir, S. J. U. Hassan, A. Mehdi, A. Abdelmaksoud, Z. Haider, G.-S. Shin, and C.-H. Kim, "A comprehensive review of vehicle-to-grid (v2g) technology as an ancillary services provider," *Results in Engineering*, p. 106813, 2025.
- [2] R. A. Jacob, S. Paul, S. Chowdhury, Y. R. Gel, and J. Zhang, "Real-time outage management in active distribution networks using reinforcement learning over graphs," *Nature Communications*, vol. 15, no. 1, p. 4766, 2024.
- [3] Y. Chen, R. A. Jacob, Y. R. Gel, J. Zhang, and H. V. Poor, "Learning power grid outages with higher-order topological neural networks," *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 720–732, 2023.
- [4] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, and J. Sun, "Two-timescale voltage control in distribution grids using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2313–2323, 2019.
- [5] R. Hossain, M. Gautam, M. MansourLakouraj, H. Livani, M. Benidris, and Y. Baghzouz, "Soft actor critic based volt-var co-optimization in active distribution grids," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2022, pp. 01–05.
- [6] S. Nematshahi, D. Shi, F. Wang, B. Yan, and A. Nair, "Deep reinforcement learning based voltage control revisited," *IET Generation, Transmission & Distribution*, vol. 17, no. 21, pp. 4826–4835, 2023.
- [7] W. Cui, J. Li, and B. Zhang, "Decentralized safe reinforcement learning for inverter-based voltage control," *Electric Power Systems Research*, vol. 211, p. 108609, 2022.
- [8] M. Ledro, L. Calearo, J. M. Zepter, T. Gabderakhmanova, and M. Marinelli, "Influence of realistic ev fleet response with power and energy controllers in an ev-wind virtual power plant," *Sustainable Energy, Grids and Networks*, vol. 31, p. 100704, 2022.
- [9] H. Xie, G. Song, Z. Shi, J. Zhang, Z. Lin, Q. Yu, H. Fu, X. Song, and H. Zhang, "Reinforcement learning for vehicle-to-grid: A review," *Advances in Applied Energy*, p. 100214, 2025.
- [10] H. D. Kaushik, J. Wang, R. A. Jacob, and J. Zhang, "Electric vehicle charger infrastructure planning: Demand estimation, coverage optimization over an integrated power grid," *arXiv preprint arXiv:2509.23699*, p. , 2025.
- [11] J. Wang, H. D. Kaushik, R. A. Jacob, and J. Zhang, "Spatiotemporal planning of electric vehicle charging infrastructure: Demand estimation and grid-aware optimization under uncertainty," *iScience*, vol. 28, no. 9, 2025.