

High-Speed Voltage Control in Active Distribution Systems with Smart Inverter Coordination and DRL

Mohammad Golgol, *Student Member, IEEE*, and Anamitra Pal, *Senior Member, IEEE*

School of Electrical, Computer and Energy Engineering

Arizona State University, Tempe, AZ, USA

mgolgol@asu.edu, anamitra.pal@asu.edu

Abstract—The increasing penetration of renewable energy resources in distribution systems necessitates high-speed monitoring and control of voltage for ensuring reliable system operation. However, existing voltage control algorithms often make simplifying assumptions in their formulation, such as *real-time availability* of smart meter measurements (for monitoring), or *real-time knowledge* of every power injection information (for control). This paper leverages the recent advances made in high-speed state estimation for real-time unobservable distribution systems to formulate a deep reinforcement learning (DRL)-based control algorithm that utilizes the state estimates *alone* to control the voltage of the entire system. The results obtained for a modified (renewable-rich) IEEE 34-node distribution feeder indicate that the proposed approach excels in monitoring and controlling voltage of active distribution systems.

Index Terms—Deep reinforcement learning, Distribution system state estimation, Inverter coordination, Voltage control

I. INTRODUCTION

The imperative to reduce carbon emissions has spurred the integration of renewable energy resources into modern power systems. However, the substantial incorporation of renewable energy, particularly at the distribution-level in the form of rooftop solar photovoltaic (PV), poses serious challenges to reliable system operation. High penetration of PVs have caused elevated voltages and voltage fluctuations in active distribution systems [1, 2]. Furthermore, conventional voltage control devices for the distribution system such as on-load tap changers (OLTCs), distribution voltage regulators (DVRs), and/or capacitor banks (CBs) [3] are not designed for frequent use (e.g., multiple times in a day). This implies that relying on them primarily for voltage control in renewable-rich distribution grids will cause increased wear-and-tear, and eventually accelerate loss-of-life of these equipment.

Due to the above-mentioned reasons, current research is focused on exploiting the capability of modern smart inverters in providing reactive power (Q) support through Volt-VAr (VV) coordination [4]. This can be done through *optimization-based* approaches or *data-driven* approaches. The former involves optimizing VV settings using deterministic or stochastic methods [5–7]. However, their reliance on model accuracy as well as lower speed of online operation have been identified as possible shortcomings. As such, data-driven approaches involving machine learning (ML), particularly deep reinforcement learning (DRL), are gaining considerable attention. Ref.

[8] cast the VV optimization problem within a deep Q-network (DQN) framework to overcome modeling limitations. Refs. [9–11] used the soft-actor critic algorithm, while [12, 13] employed the deep deterministic policy gradient (DDPG) to solve the voltage control problem.

However, one or more of the following issues were identified with the DRL-based control approaches developed in [8–13]. Most of them assumed access to *full system information* such as power consumption, solar power generation, and voltage, *in real-time*. In practice, such data is not readily available. For instance, smart meters measure voltage and *average* power at intervals of at least 15 minutes [14]. More importantly, these measurements are not reported to the utility until *after a few hours* [15]. This means that they cannot be used directly for performing distribution system state estimation (DSSE), which was the basis for these control architectures. Furthermore, these approaches assumed that conventional distribution voltage control devices (OLTCs, DVRs, CBs), are designed for frequent operation, which is not the case in reality.

The control approach proposed in this paper advances the state-of-the-art in the following ways: (1) It sidesteps the need for full system information in real-time by relying solely on a state estimator that uses measurements *at the feeder-head alone* to perform high-speed DSSE. (2) Its training and performance verification are done using this state estimator. (3) It only changes settings of the smart inverters using remote signals for mitigating voltage violations; i.e., it does not change settings of the conventional control devices. (4) It causes minimal active power curtailment. The results obtained using the IEEE 34-node *unbalanced* distribution feeder proves that the proposed innovative, infrastructure-light approach enables the monitoring and management of active distribution systems that have high penetration of rooftop solar PV.

II. ML-BASED TIME SYNCHRONIZED DSSE

In a modern distribution system, sensors are usually placed at the root node (e.g., SCADA, μ PMU) or the leaf nodes (e.g., smart meters, solar meters). Consequently, the main trunk of the system remains concealed from power utilities from monitoring and control perspectives. Moreover, the high latency period of smart meters prevents their data from being reported in real-time. This limited real-time visibility lowers the effectiveness of conventional approaches for DSSE [15].

To overcome this challenge and achieve comprehensive situational awareness with minimal sensor coverage, an ML-

This work was supported in part by the National Science Foundation (NSF) under the grant ECCS-2145063.

based high-speed state estimator was developed in [16]. It used Bayesian inference to establish a mapping relationship between μ PMU measurements and state variables (voltage magnitude and angle) of all network nodes. Historical data from slow-timescale, high-latency sensors were used for training, while fast-timescale data was used for testing. In this paper, we employ this ML-based state estimator to perform end-to-end high-speed DSSE using μ PMUs placed *only at the feeder-head*. The DSSE outputs are used to (a) train the proposed DRL-based control, and (b) test the performance of the control. More details about the offline training and online implementation of this state estimator can be found in [17, 18].

III. DRL-BASED INVERTER CONTROL

A. Modeling voltage control as an optimization problem

We start by formulating the voltage control problem as a constrained optimization problem, as shown below:

$$\min_{Q_{PV}} J = \sum_{t=1}^T \left(\sum_{n=1}^N |V[n,t] - V_{\text{nominal}}| \right) \quad (1)$$

subject to:

$$\begin{aligned} & V_R[n,t] \left(\sum_{m=1}^N (G_{n,m} V_R[m,t] - B_{n,m} V_I[m,t]) \right) \\ & + V_I[n,t] \left(\sum_{m=1}^N (G_{n,m} V_I[m,t] + B_{n,m} V_R[m,t]) \right) \\ & + P[n,t] = 0 \quad \forall n \in N \text{ and } \forall t \in T \end{aligned} \quad (2)$$

$$\begin{aligned} P[n,t] &= P_{\text{load}}[n,t] - P_{PV}[k,t] - P_F[t] \\ &\quad \forall n \in N, t \in T, \text{ and } k \in K \end{aligned} \quad (3)$$

$$0 \leq P_{PV}[k,t] \leq P_{\text{rated}}[k] \quad \forall k \in K, t \in T \quad (4)$$

$$\begin{aligned} & V_I[n,t] \left(\sum_{m=1}^N (G_{n,m} V_R[m,t] - B_{n,m} V_I[m,t]) \right) \\ & - V_R[n,t] \left(\sum_{m=1}^N (G_{n,m} V_I[m,t] + B_{n,m} V_R[m,t]) \right) \\ & + Q[n,t] = 0 \quad \forall n \in N \text{ and } \forall t \in T \end{aligned} \quad (5)$$

$$\begin{aligned} Q[n,t] &= Q_{\text{load}}[n,t] - Q_{PV}[k,t] - Q_F[t] \\ &\quad \forall n \in N, t \in T, \text{ and } k \in K \end{aligned} \quad (6)$$

$$-Q_{\text{rated}}[k] \leq Q_{PV}[k,t] \leq Q_{\text{rated}}[k] \quad \forall k \in K, t \in T \quad (7)$$

$$V_{\min} \leq V[n,t] \leq V_{\max} \quad \forall n \in N, t \in T \quad (8)$$

The objective function (1) minimizes the absolute difference between the voltage of node n at time t , denoted by $V[n,t]$, and the nominal rated voltage, denoted by V_{nominal} (set at 1 p.u. by default) for all n and t . This objective is attained by adjusting the Q output of smart inverters. The constraints that must be satisfied are defined in (2)-(8), in which $V_R[n,t]$ and $V_I[n,t]$ denote the real and imaginary components of $V[n,t]$, $G_{n,m}$ and $B_{n,m}$ denote the conductance and susceptance between nodes n and m , P_F and Q_F denote the active and reactive power drawn from the feeder-head, and K , N ,

and T denote sets of PV sources, all nodes' indices, and time-steps, respectively. Equations (2)-(6) ensure active and reactive power balance. In (7), $Q_{\text{rated}}[k]$ refers to the maximum available reactive power that could be provided by the k^{th} smart inverter, while $S_{\text{rated}}[k]$ and $P_{\text{rated}}[k]$ are the rated values of apparent and active power for that inverter, respectively. The desired voltage limits are introduced in (8).

In order to solve (1)-(8) analytically, one needs *instantaneous values* of the power injections, which are not readily available in most distribution systems. Furthermore, solving this problem at each time-step can be time-consuming, particularly as the network size increases. To address these issues, we propose a data-driven approach based on DRL that employs the DDPG algorithm. The proposed approach eliminates the need for precise model parameters and allows the agent to adapt to real-time data. Moreover, the trained agent can be used in real-time without significant response delays. In the following sub-section, we detail how this constrained optimization problem can be transformed into a Markov decision process (MDP), which is the prerequisite for solving any problem using DRL.

B. Formulating voltage control as an MDP problem

An MDP is a mathematical model for decision-making, involving an agent interacting with an environment over discrete time-steps. Its key components are state, action, environment, and reward. The state represents the system's different configurations. In our case, we use voltage magnitudes of every node, as it encapsulates all the necessary information required to make an informed decision. The following equation specifies the system state, s , at time t :

$$s_t = (|V[n,t]|), \quad \forall n \in N \quad (9)$$

Actions are the choices available to the DRL agent. For our problem, they are described by the following equations:

$$Q_{PV}[k,t] = a_t^l \cdot Q_{\text{rated}}[k] \quad \text{where } -1 \leq a_t^l \leq 1 \quad (10)$$

$$A_t = (a_t^l, \quad \forall l \in L) \text{ at each time } t \quad (11)$$

In (10), each agent gives a specific coefficient a_t^l that determines how much available reactive power, of those smart inverters that are controlled by it, should be utilized for voltage control. L is the set of all agents, with (11) specifying the action of all the agents at time t . The environment models the system's dynamics, determining the next state based on the current state-action pair. The reward function defines the immediate benefit or cost of taking the action. For our objective of maintaining voltage within desired range with minimum active power curtailment, the following reward function is proposed:

$$r_t = - \left(\lambda \sum_{n \in N} \Lambda(V[n,t]) + \eta \sum_{k \in K} \Gamma(Q_{PV}[k,t]) \right) \quad (12)$$

$$\Lambda(V) = \begin{cases} \|V - V_{\text{nominal}}\|^2 & V_{\min} \leq V \leq V_{\max} \\ |V - V_{\text{nominal}}| & \text{otherwise} \end{cases} \quad (13)$$

$$\Gamma(q_k) = \begin{cases} 0 & 0 \leq |q_k| \leq Q_{\max}[k] \\ ||q_k| - Q_{\max}[k]| & \text{otherwise} \end{cases} \quad (14)$$

$$Q_{\max}[k] = \sqrt{S_{\text{rated}}^2 - P_{\text{PV}}^2[k]} \quad (15)$$

where, $\Lambda(\cdot)$ corresponds to the voltage control effort, $\Gamma(\cdot)$ corresponds to the power curtailment effort, and $Q_{\max}$ refers to the maximum reactive power that can be supplied without active power curtailment [19]. As the cardinality of N and K may differ, we introduce λ and η to balance the two efforts. Fine-tuning these parameters enables us to align the network's behavior with operational guidelines.

C. Control execution using DDPG algorithm

The DDPG algorithm is well-suited for continuous action spaces [20], which fits the scope of our problem. It consists of an actor and a critic network, which are neural networks parameterized by θ_π and θ_ϑ , respectively, as shown below:

$$A_t = \pi(s_t | \theta_\pi) \quad (16)$$

$$\vartheta(s_t, A_t | \theta_\vartheta) \approx r_t + \gamma \vartheta(s_{t+1}, \pi(s_{t+1} | \theta'_\pi) | \theta'_\vartheta) \quad (17)$$

The actor (16) maps states to actions based on the policy function, while the critic evaluates the action (17). The discount factor $\gamma \in [0, 1]$, governs the consideration of future rewards. Note that to ensure robustness and stability during training, target networks (θ'_π and θ'_ϑ) are employed for both the actor and the critic, respectively, which are incrementally updated to shadow the learned networks. The training utilizes a deterministic policy gradient for the actor and a temporal difference (TD) error for the critic.

1) *Offline Stage*: The offline stage, depicted in Fig. 1, is the training phase, where the agent is exposed to various power system scenarios, represented by (P, Q, P_{PV}) , to learn the optimal control actions. Each scenario is processed through the environment and DSSE modules to acquire the state, s_t . Subsequently, the agent determines the action A_t using (16), leading to the new state s_{t+1} . The immediate reward r_t is computed given the transition from s_t to s_{t+1} . These experiences are stored as transitions in the replay buffer until the end of each episode, and a new scenario is then selected for the next episode. When the replay buffer is filled, a batch of transitions is utilized to adjust the weights of the actor and critic networks iteratively until the TD error converges. This is explained mathematically in the following equations:

$$\text{TD} = \frac{1}{B} \sum_{i=1}^B (\vartheta(s_i, A_i | \theta_\vartheta) - (r_i + \gamma \vartheta(s_{i+1}, \pi(s_{i+1} | \theta'_\pi) | \theta'_\vartheta))) \quad (18)$$

$$\nabla_{\theta_\pi} L \approx \frac{1}{B} \sum_{i=1}^B \nabla_A \vartheta(s, A | \theta_\vartheta) |_{s_i, A=\pi(s_i)} \nabla_{\theta_\pi} \pi(s | \theta_\pi) |_{s_i} \quad (19)$$

$$\begin{cases} \theta'_\vartheta \leftarrow \tau \theta_\vartheta + (1 - \tau) \theta'_\vartheta \\ \theta'_\pi \leftarrow \tau \theta_\pi + (1 - \tau) \theta'_\pi \end{cases} \quad (20)$$

The update process for the critic network involves minimizing (18), while the actor network undergoes updates

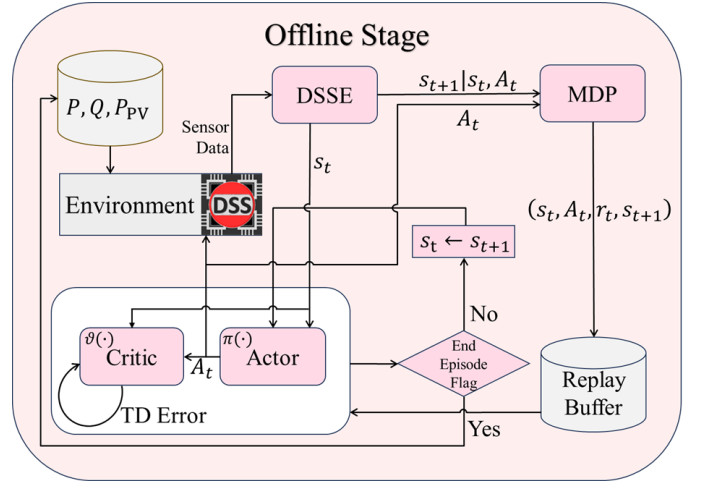


Fig. 1: Offline training process of proposed control

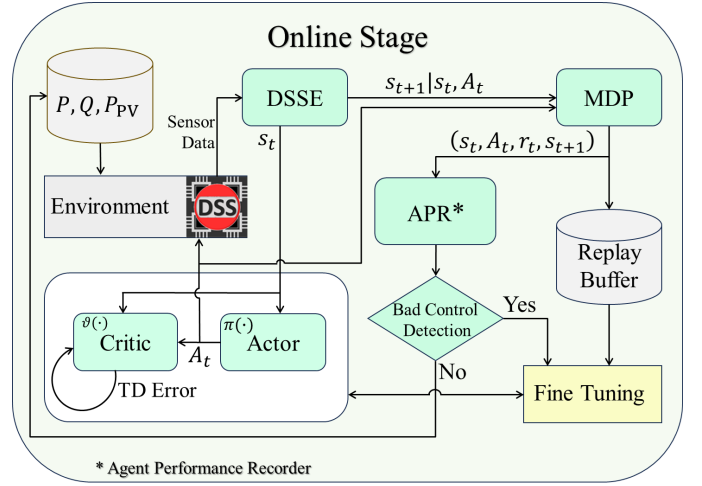


Fig. 2: Online execution of proposed control

through the sampled policy gradient method outlined in (19). Additionally, the parameters of the target networks are updated according to the specifications in (20). The term B in (18)-(19) represents the number of transitions in a batch, while τ in (20) is the soft update coefficient.

2) *Online Stage*: In the online phase, shown in Fig. 2, the trained agent is subjected to testing. The DSSE module processes real-time sensor data to estimate the states. The voltage magnitudes obtained at the output of the DSSE module serve as the current state s_t for the agent. The agent then determines the appropriate action A_t to adjust the system's state. The resulting state s_{t+1} and reward r_t are determined through the DSSE and MDP modules, respectively. The agent performance recorder (APR) continuously monitors the agent's decision-making process (via the DSSE and MDP modules). If the APR identifies poor control performance, it initiates a fine-tuning process for the agent's parameters, enabling it to adapt to the evolving operational conditions of the system.

IV. SIMULATION RESULT

A. Case-study and data preparation

We conducted simulations on the IEEE 34-node unbalanced feeder to assess the impact of high PV penetration at the distribution level. We added PV systems at all load locations, with PV power generation ranging from 74% to 105% of the respective load. As this system lacks historical smart meter data, we generated diverse scenarios using Pecan Street data [21]. This dataset contains smart meter load and rooftop PV data from 11 houses. We selected data corresponding to July with a time interval of 12 Noon to 1:00 PM, representing the period of maximum PV power generation. Finally, since the 34-node feeder is at the primary voltage level, we randomly combined household level load and PV profiles from Pecan Street to create aggregated profiles at the primary level.

B. Parameter settings

The DNN model in our ML-based DSSE featured a sequential architecture with 5 hidden layers, each consisting of 200 neurons having ReLU activation function. To prevent overfitting, a 50% dropout rate and batch normalization were applied after each hidden layer. The DNN, with an input size of 12 (voltage and current phasor values of all three phases at the feeder-head) and output of 270 (voltage magnitudes and angles of all phases of every node), was trained for 100 epochs using the mean squared error loss function. The optimizer used was Adam and the learning rate was 0.095.

The actor and critic networks in our DDPG algorithm utilized two hidden layers with 400 and 300 neurons each, a learning rate of 0.001, and a batch size of 64. The activation functions in the critic network were ReLU for both layers, while it was ReLU for the first layer and tanh for the second layer in the actor network. The soft update coefficient, denoted by τ , was fixed at 0.001, while the exploration-exploitation parameters, μ and σ , were initially set to 0.1 and 0.5, respectively. As the training progressed, σ gradually decreased from 0.5 to 0.005. The input layer of the actor network had a size of 135, which corresponded to the number of phase voltage magnitudes obtained from DSSE. The critic network's input layer had a size of $135 + 1$, combining state and action information. Both networks had an output layer size of 1. For the reward function, the results were achieved with values of 1 for λ and 0.5 for η . The replay buffer size was set to 10,000.

C. Performance of ML-based DSSE

The DSSE results for the modified (renewable-rich) IEEE 34-node system with one μ PMU at the feeder-head is shown in Table I. The mean absolute percentage error (MAPE) metric is used for the magnitudes, while the mean absolute error (MAE) metric is used for the angles. The results were obtained with 1% Gaussian noise in the μ PMU data, and confirm the ability of ML to overcome the real-time unobservability issue in distribution systems. This state estimator is now used to train and test the proposed control.

TABLE I: DSSE results for modified IEEE 34-node system

	Phase A	Phase B	Phase C
Magnitude MAPE (%)	0.0804	0.0957	0.0228
Phase MAE (°)	0.0248	0.0228	0.0212

D. Assessment of voltage regulation performance

Fig. 3 shows the average cumulative reward across 100 training runs, with the solid line representing the mean reward at each episode and the shaded area indicating the variability in the performance of the agent (captured via standard deviation of the reward). The initial decline in the reward is due to early exploration and learning. Over the course of 100 episodes, we observe an overall upward trend in the reward, suggesting that the agent is effectively learning and improving its policy. The fluctuations in the reward trajectory are due to the exploratory nature of DRL and the stochastic elements within the environment. The diminishing width of the shaded region over time reflects a reduction in performance variance, pointing to the agent's increasing stability and the algorithm's convergence towards a consistent policy.

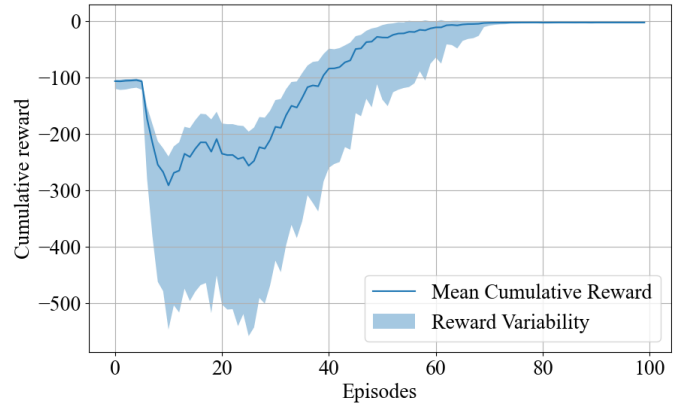


Fig. 3: Evolution of cumulative reward over 100 training runs

Fig. 4 shows the Phase A voltage magnitude results obtained in the absence of control (i.e., inverters operating at unity power factor) across 10,000 scenarios. It can be observed from the figure that the mean voltage (solid blue line) and its variance (shaded area) frequently exceed the upper voltage limit (red dotted line). The several instances of over-voltage in the different nodes indicate the potential risks and challenges associated with managing voltage in active distribution grids. Similar observations were made for the other phases as well.

After implementing the proposed control algorithm for the same scenarios, the resulting voltage (mean and variance) consistently remained within the acceptable range (see Fig. 5). The absence of values beyond the designated limits for any of the nodes reflects the algorithm's capability to avert over-voltage (and under-voltage) conditions across the distribution system. This result highlights its practical utility in maintaining the operational voltage within desirable limits even in presence of considerable amounts of renewable generation.

The overall process of feeding μ PMU data into ML-based DSSE, and using its outputs to determine the appropriate control action, took less than 1 second. This substantiates the

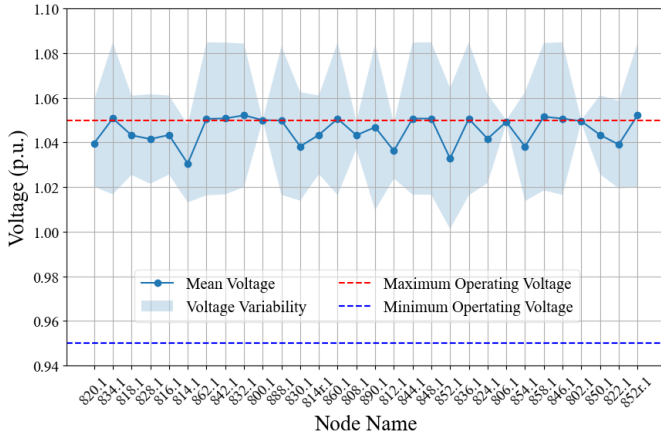


Fig. 4: Phase A voltage profile at unity power factor operation

real-time applicability claim of our proposed method (see last paragraph of Section III-A). Moreover, by learning via system interaction, it circumvents dependency on model parameters.

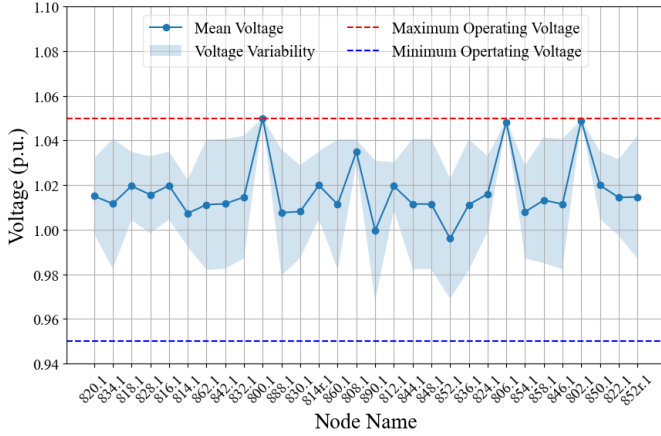


Fig. 5: Phase A voltage profile with proposed coordination

V. CONCLUSION

The increasing penetration of renewable energy resources in distribution systems necessitates the development and implementation of advanced voltage control strategies. Prior approaches for voltage control were often limited by heavy online computational burden and/or requirement of an extensive sensing and communication system. The proposed DRL-based control strategy relies on estimates obtained from an ML-based state estimator that provides high-speed outputs for real-time unobservable systems. The results indicate that the proposed infrastructure-light control effectively maintains voltage across the distribution network, even with the inherent variability introduced by renewable energy resources. This research underscores the potential for smart grid technologies to evolve towards more resilient and adaptive management systems in the face of increasing renewable energy utilization.

REFERENCES

[1] A. Amanipoor, M. S. Golsorkhi, N. Bayati, and M. Savaghebi, “V-Iq based control scheme for mitigation of transient overvoltage in

distribution feeders with high PV penetration,” *IEEE Transactions on Sustainable Energy*, vol. 14, no. 1, pp. 283–296, 2022.

[2] S. Singh, S. Veda, S. P. Singh, R. Jain, and M. Baggu, “Event-driven predictive approach for real-time Volt/VAR control with CVR in solar PV rich active distribution network,” *IEEE Transactions on Power Systems*, vol. 36, no. 5, pp. 3849–3864, 2021.

[3] A. Suresh, R. Bisht, and S. Kamalasadan, “A coordinated control architecture with inverter-based resources and legacy controllers of power distribution system for voltage profile balance,” *IEEE Transactions on Industry Applications*, vol. 58, no. 5, pp. 6701–6712, 2022.

[4] Y. Li and J. Yan, “Cybersecurity of smart inverters in the smart grid: A survey,” *IEEE Transactions on Power Electronics*, 2022.

[5] Z. Tang, D. J. Hill, and T. Liu, “Distributed coordinated reactive power control for voltage regulation in distribution networks,” *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 312–323, 2020.

[6] C. Zhang, Y. Xu, Y. Wang, Z. Y. Dong, and R. Zhang, “Three-stage hierarchically-coordinated voltage/VAR control based on PV inverters considering distribution network voltage stability,” *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 868–881, 2021.

[7] D. Dalal, M. Sondharangalla, R. Ayyanar, and A. Pal, “Improving photovoltaic hosting capacity of distribution networks with coordinated inverter control – a case study of the EPRI J1 feeder,” 2023. [Online]. Available: <https://arxiv.org/abs/2311.02793>

[8] Y. Zhang, X. Wang, J. Wang, and Y. Zhang, “Deep reinforcement learning based Volt-VAR optimization in smart distribution systems,” *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 361–371, 2020.

[9] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, Z. Chen, and F. Blaabjerg, “Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs,” *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4137–4150, 2021.

[10] J. Wang, W. Xu, Y. Gu, W. Song, and T. C. Green, “Multi-agent reinforcement learning for active voltage control on power distribution networks,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 3271–3284, 2021.

[11] D. Cao, J. Zhao, J. Hu, Y. Pei, Q. Huang, Z. Chen, and W. Hu, “Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control,” *IEEE Transactions on Smart Grid*, 2023.

[12] X. Zhang, Y. Liu, J. Duan, G. Qiu, T. Liu, and J. Liu, “DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources,” *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5465–5475, 2021.

[13] H. Liu and W. Wu, “Online multi-agent reinforcement learning for decentralized inverter-based Volt-VAR control,” *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 2980–2990, 2021.

[14] A. Lang, Y. Wang, C. Feng, E. Stai, and G. Hug, “Data aggregation point placement for smart meters in the smart grid,” *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 541–554, 2021.

[15] B. Azimian, R. S. Biswas, A. Pal, and L. Tong, “Time synchronized state estimation for incompletely observed distribution systems using deep learning considering realistic measurement noise,” in *2021 IEEE Power & Energy Society General Meeting (PESGM)*, 2021, pp. 1–5.

[16] B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarthy, “State and topology estimation for unobservable distribution systems using deep neural networks,” *IEEE transactions on instrumentation and measurement*, vol. 71, pp. 1–14, 2022.

[17] B. Azimian, S. Moshtagh, A. Pal, and S. Ma, “Analytical verification of deep neural network performance for time-synchronized distribution system state estimation,” *Journal of Modern Power Systems and Clean Energy*, p. to appear, 2023.

[18] B. Azimian, R. S. Biswas, and A. Pal, “Application of AI and machine learning algorithms in power system state estimation,” in *Cyber-physical power systems: challenges and solutions by AI/ML, big data, blockchain, IoT, and information theory paradigms*. Wiley-IEEE Press, 2024.

[19] “IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces,” *IEEE Std 1547-2018*, pp. 1–138, 2018.

[20] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[21] Pecan Street Research Institute, “The Dataport Database.” [Online]. Available: <https://pecanstreet.org/dataport/>