Cooperative Multi-Agent Reinforcement Learning Method for Fast Voltage Regulation using Distributed Wind Turbine Generators

Miguel Jimenez-Aparicio
Sandia National Laboratories
Albuquerque, NM, USA

Rachid Darbali-Zamora
Sandia National Laboratories
Albuquerque, NM, USA

2023 IEEE PES Innovative Smart Grid Technologies Latin America (ISGT-LA) | 979-8-3503-3696-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISGT-LA56058.2023.10328322

***Abstract*—This paper proposes a Multi-Agent Reinforcement Learning (MARL) voltage regulation method for distribution systems with presence of Wind Turbine Generators (WTGs). The control employs reactive power injection or absorption in three WTGs to regulate feeder voltages. A weighted global voltage metric is calculated to summarize nodes' voltages according to their relevance. Three independent RL Agents are trained using policy parameter sharing. Models are trained and tested on real-time Hardware-In-the-Loop setup. The proposed control effectively regulates voltage through the system, bringing median and mean voltages closer to 1 p.u. and reducing standard deviation.**

***Index Terms*—Voltage Regulation, DERMS, Wind Turbine Generator, Multi-Agent System, Reinforcement Learning**

I. Introduction

The ongoing deployment of Distributed Energy Resources (DERs) into the distribution grid has multiple advantages, such as losses minimization, and a more flexible and resilient power grid. However, this is not exempt from risks. Distribution systems subjected to high penetration of Variable Renewable Energy (VRE) can experience sudden voltage swings. For example, solar Photovoltaic (PV) output can be reduced up to 15% per second due to clouds [1]. Many companies are developing DERMS for DERs coordinated management. According to the IEEE 2030.11-2021 standard, DERMS basic functionalities include: DER registration, dispatch/scheduling, and monitoring. In addition, most commercial DERMS include grid-support functionalities, such as voltage or frequency regulation.

"This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan."

The authors would like to acknowledge the financial support provided by the U.S. Department of Energy (DOE), Office of Energy Efficiency (EERE), Wind Energy Technology Office (WETO) for the Wind Hybrid Integration Platform (WHIP) project under Work Breakdown Structure No. 1.2.6.701 as well as the contributions from Bret Barker and Patrick Gilman.

The IEEE 1547-2018 standard encourages the use of DERs on grid-support activities. The fast dynamics of DER inverters allow for a significantly faster and less intrusive response than conventional voltage regulation devices, such as on-line tap-changing (OLTC) transformers or capacitor banks. Most DERMS solutions and research focus on Energy Storage Systems (ESS) and PV inverters. Unfortunately, the potential of distributed wind turbines is overlooked, even though they can be successful grid-support functions providers [2].

The contribution of this paper is the design, training and validation in Hardware-In-the-Loop (HIL) of a method for fast voltage regulation (dispatch interval of 1 second), by means of reactive power control, using multiple Wind Turbine Generators (WTGs) simultaneously, and based on Multi-Agent Reinforcement Learning (MARL) with parameter-sharing.

II. Background

The basic properties of Multi-Agent Systems (MAS) on power system environments, which are outlined in [3], are:

- An agent is an entity that is capable of interacting with an external environment.
- An agent is capable of autonomous reasoning, independently on the number of agents and the environment.
- An agent has partial/fully environment observability.
- An agent can communicate with other agents.

MAS have been researched on multiple power system applications: restoration, network controls, microgrid profit maximization, etc. Regarding voltage regulation on distribution systems, a few papers propose optimization methods for minimizing voltage deviations. The work in [4] relies on the coordinated management of OLTC transformers. In [5], the voltage regulation is based on the power factor adjustment of PV inverters. The control acts every few seconds. The work in [6] tests in a Power-Hardware-in-the-Loop (PHIL) setting a reactive power optimization algorithm for WTGs. Their setups counts with a WTG emulator, real PV and a super capacitor.

More recently, the application of RL for optimizing grid support functions in on the rise. The work in [7] designs a Soft Actor-Critic (SAC) agent to avoid voltage deviations and minimize active power curtailment on PVs while being aware of different system topologies. In [8], a reactive power

Authorized licensed use limited to: Sandia National Laboratories. Downloaded on November 05,2025 at 19:49:16 UTC from IEEE Xplore. Restrictions apply.

controller based on a Deep Deterministic Policy Gradient (DDPG) is trained for fast voltage regulation using a WTG. The action time step is just of 50 milliseconds. In order to achieve cooperation among devices, MARL approaches have been also explored. In [9], the reactive power setting of each PV inverter is determined by its respective agent. The agents have only local system observability, although the reward is calculated based on the system-wide voltage deviation. The control action is delivered every hour. A Multi-Agent SAC (MASAC) is proposed in [10] for hybrid systems (PVs and WTGs). The dispatch interval is 5 minutes. In [11], a mixed-integer optimization method and MARL are combined to reduce power loss and voltage deviations. The state for each agent correspond to the local nodal observations. The reward can be calculated either locally or globally. The time-step for the control signal is 5 minutes.

Furthermore, the work in [12] provide insightful information on the experimental HIL demonstration setup. The real-time measurements on the network can be extracted using Modbus from the OPAL-RT machine (Modbus slave). On the other hand, the voltage regulation algorithm can be executed on an external PC (Modbus master), which has the capability of writing setpoints in the simulation.

III. Methodology

A. The System

The model hosted in our OPAL-RT HIL capabilities is a digital-twin of the Scaled Wind Farm Technology (SWiFT) test site located in Lubbock, Texas [13]. The system is modeled in MATLAB/Simulink. A representation of the system can be seen in Fig. 1. This distribution system consists of 10 buses. The nominal voltage of buses 1 to 9 is 12.47 kV, while Bus 10 is rated at 480 V. The system has 3 identical WTGs, 200 kW each, located at Buses 6 to 8. These devices are capable of providing grid-support functions through adjustable active and reactive power setpoints [14]. These WTGs are rated at 480V and interfaced to the medium voltage grid through a step-up transformer. A 500 kW variable load and a 150 kW solar PV are connected to Bus 10. For this purpose, a step-down transformer is located between Bus 9 and 10. In this work, it is considered that the 3 WTGs work with active power priority. The solar PV does not provide any grid support function.

B. The Controller

The proposed voltage regulation grid-support functionality is modeled as a MARL problem. The 3 WTGs are considered as agents that cooperate on the voltage regulation task. The 3 RL agents are based on Deep Deterministic Policy Gradient (DDPG) models. DDPG is an actor-critic RL model in which the actor network creates a policy to map an optimal action with the received environment observations. The critic network estimates the expectation of the long-term reward for such action. Both networks are trained simultaneously. All agents have the same structure, which is shown in Fig. 2. The networks are composed of Fully Connected layers (FC) with the indicated input dimensions and non-linear activation layers.

The MARL training is based on the concept of "parameter sharing": each WTG is controlled by an independent RL agent, but all RL agents share the same policy network. Parameter sharing on MARL problems have demonstrated to increase training stability on cooperative tasks and learn optimal policies [15]. In this case, the 3 WTGs are Inverter-Based Resources (IBRs) with identical capabilities and action spaces, and similar observation spaces. Therefore, training a

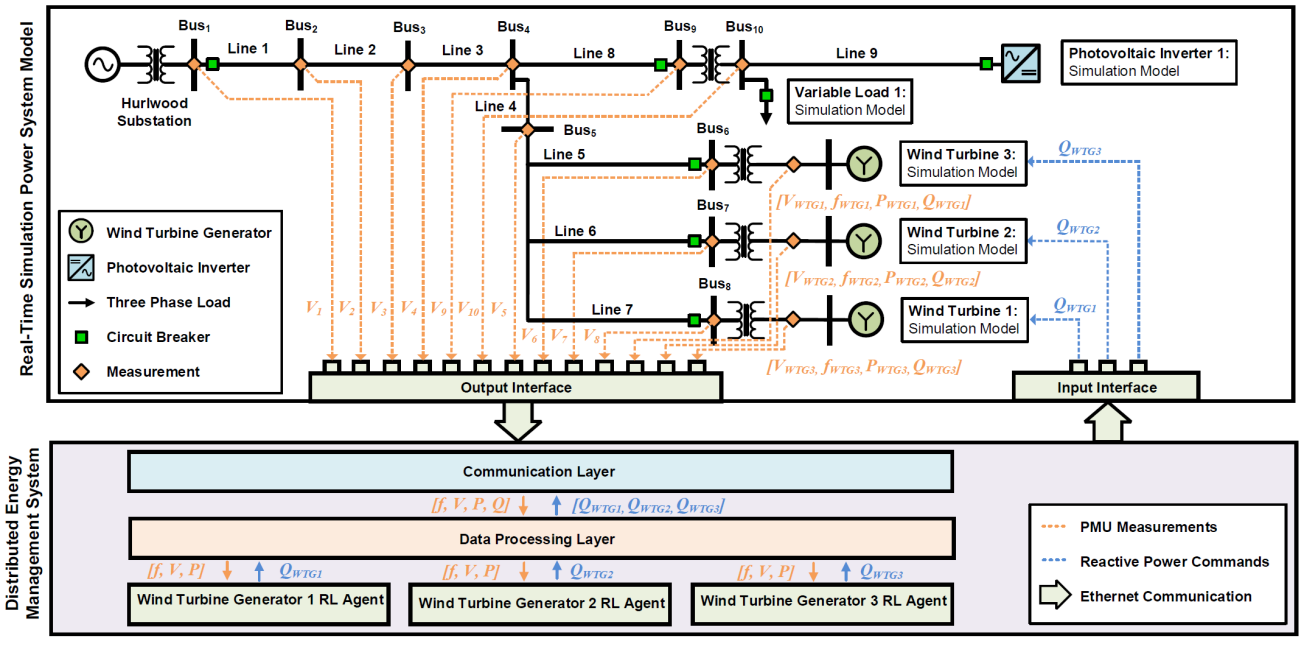

Fig. 1. System voltages comparison with and without the proposed voltage regulation controller.

Authorized licensed use limited to: Sandia National Laboratories. Downloaded on November 05,2025 at 19:49:16 UTC from IEEE Xplore. Restrictions apply.

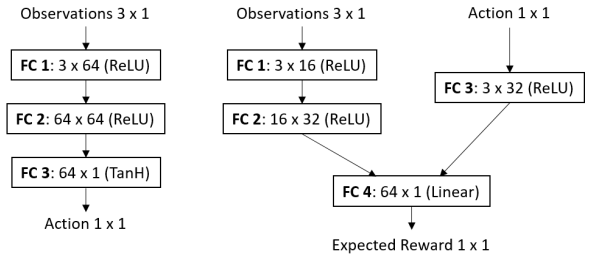

Fig. 2. Actor (left) and Critic (right) network structures.

single neural network with joint experiences of the 3 WTGs ensures an optimal response for each of them. Furthermore, obtaining a single policy that is optimal for multiple devices is advantageous for the method' scalability: If an additional IBR is added, the same controller could be employed.

The WTG agents interact with the environment by means of injecting or absorbing reactive power. Voltage regulation through reactive power exchange is fast, efficient, and doesn't require the deployment of additional equipment [8]. The WTG agents have partial environment observability. Each RL Agent i has access to the following measurements: RMS voltage at the Point-Of-Interconnection (POI) $V_{RMS,i}$, frequency f_i, and active and reactive power injections, P_i and Q_i. Furthermore, the agents have access to a global voltage metric $V_{Global,t}$ that efficiently summarizes the voltage across the systems for each time-step t. An agent can infer other agents' actions by observing this metric. The proposed $V_{Global,t}$ metric consists of a weighted average of the bus voltages $V_{b,t}$ in the system, where b refers to the bus index. Weights K_b are assigned based on nodes' criticality, and are gathered in Table I.

TABLE I
NODE'S WEIGHT FOR AVERAGED GLOBAL VOLTAGE METRIC

Weight K_b	Buses b
5	10, WTG 1, WTG 2, WTG 3
2	6, 7, 8, and 9
1	1, 2, 3, 4, and 5

The larger weight, 5, is assigned to nodes where IBRs and loads are connected. According to the ANSI C.84.1 standard, buses with a nominal voltage lower than 600V must be regulated in a $\pm$ 5% interval. This is the case of the buses where the PV, WTGs and the load are connected. The IEEE 1547-2018 standard indicates that IBRs' normal voltage range is between 0.88 p.u. and 1.1 p.u. Measurements from the high-voltage side of transformers where WTGs and the PV are connected receive a medium weight, 2. Lastly, all other buses receive a low weight of 1. $V_{Global,t}$ is defined as:

$$V_{Global,t} = \frac{\sum_b K_b \times V_{b,t}}{\sum_b K_b} \ . \tag{1}$$

The instantaneous reward function for each RL Agent i, $\text{Reward}_{i,t}$, has both a global and specific term. The first one penalizes voltage deviation on the previously defined $V_{Global,t}$. The second one penalizes the control action to avoid excessive usage. The agent-specific reward is calculated as:

$$\text{Reward}_{i,t} = \text{Reward}_{V_{global},t} + \text{Reward}_{Q_i,t} \ . \tag{2}$$

The penalization function for $V_{Global,t}$, $\text{Reward}_{V_{global},t}$, is designed to have increasingly steeper slopes for larger voltage deviations, as shown in Fig. 3. The $\text{Reward}_{V_{global},t}$ penalization is calculated with the following parabolic function:

$$\text{Reward}_{V_{global},t} = 3523 \cdot V_{Global,t}^2 - 7048 \cdot V_{Global,t} \ + 3524. \tag{3}$$

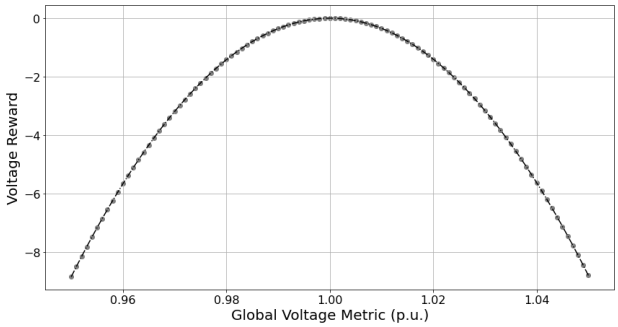

Fig. 3. Voltage reward matching.

The control action penalization for RL Agent i is calculated as shown in Eq. (4), where $\alpha = 0.15$:

$$\text{Reward}_{Q_i,t} = \alpha * Q_{i,t} \ . \tag{4}$$

C. The Training Setup

The control policy is trained on a custom HIL setup. The TensorFlow DDPG model is located on a PC (Modbus master) which communicates with a running real-time simulation on an OPAL-RT machine (Modbus slave). The training procedure has full control of the OPAL-RT machine, being able to load scenarios, retrieve measurements, and write new control actions through the Input/Output (I/O) interface. This setup is represented in Fig. 1. As mentioned earlier, the WTGs measure RMS voltage at the POI, frequency, and local active and reactive power injection. In the other nodes, a PMU measures RMS voltage. Every 1 second, these measurements are sent through Modbus protocol to the master PC, where they are received by the Communication Layer (CL) interface of the proposed solution. The voltage measurements are employed to calculate $V_{Global,t}$ on the Data Processing Layer (DPL). The $V_{Global,t}$ metric, $P_{i,t}$ and $f_{i,t}$ are passed along from the DPL to each RL Agent i for computing the next control action $Q_{i,t+1}$. The 3 control actions are looped back simultaneously once calculated.

The agents are trained using real solar irradiation, and wind-speed data. The authors have access to 62 hours of data. The load data is synthetically generated through a random walk between 0 and 500 kW. During training, solar, wind,

Authorized licensed use limited to: Sandia National Laboratories. Downloaded on November 05,2025 at 19:49:16 UTC from IEEE Xplore. Restrictions apply.

and load data is written every 1 second into the real-time simulation using the I/O interface. The 3 WTG RL Agents are trained using 600 episodes of length 100 seconds each. The testing dataset consists of 100 episodes of length 100 seconds each.

IV. Results

The episodic and average episodic rewards can be seen in Fig. 4. After episode 500, the average episodic reward reaches a plateau, which indicates that the models are fully trained.

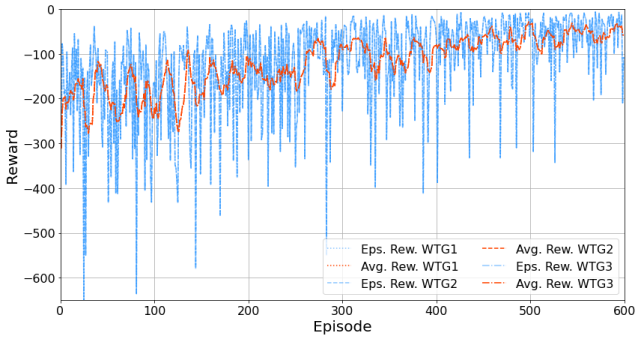

Fig. 4. Episodic rewards during training.

The recorded results over the 100 episodes test dataset show the performance of the proposed method. In Table II, the median, mean voltages and the standard deviation (Std. Dev.) at the WTGs' POI and at Bus 10 (where the PV and the load are connected to) are gathered. The cases with the proposed voltage regulation method and the base case without any control ($Q_i = 0$ for every RL Agent i) are compared. As it can be seen, the median and mean voltages at the WTGs' POI decrease in about 0.01 p.u. The voltage standard deviation decreases approximately 0.002 p.u. Regarding Bus 10, the median and mean voltage are reduced in approximately 0.01 p.u. The standard deviation at Bus 10 just decreases in 0.001 p.u. in this case. This relatively smaller reduction in standard deviation is due to the fact that the agents are far away from Bus 10, and the effect of the control action is diminished. In addition, the standard deviation at Bus 10 is significantly larger than on the WTG buses. Voltage regulation at Bus 10 could be significantly improved if the PV provided grid-support functions.

In order to provide a deeper analysis of the proposed control benefits, the recorded WTGs' RMS voltages at the POI $V_{RMS,i}$ are depicted against the injected active power P_i in Fig. 5, without and with control. In the base case, WTGs' voltages tend to increase with the larger injection of active power into the system. For a significant number of cases with a large WTG active power generation (> 0.6 p.u.), WTGs' voltages are above 1.05 p.u. On the other hand, the proposed controller is able to shift voltages down in the entire active power generation range.

In Fig. 6, it can be seen that the learned policy typically commands to absorb reactive power to drag system voltages

TABLE II
Voltage Metrics (in p.u.) for Controller and Base Cases

		V_{Bus10}	V_{WTG1}	V_{WTG2}	V_{WTG3}
Base	Median	1.012	1.029	1.029	1.029
	Mean	1.008	1.032	1.032	1.032
	Std. Dev.	0.027	0.011	0.011	0.011
DDPG	Median	**1.000**	**1.017**	**1.017**	**1.017**
	Mean	**0.996**	**1.020**	**1.020**	**1.020**
	Std. Dev.	**0.026**	**0.009**	**0.009**	**0.009**

down. For low levels of active power injection (< 0.6 p.u.), the commanded reactive power setpoints are between -0.2 p.u. and -0.4 p.u. A small decrease in WTG's POI voltage can be observed, as there is a balance between the penalization for the control effort and the observed voltage deviation. However, for high levels of active power injection (> 0.6 p.u.), the learned command is to significantly absorb reactive power between -0.6 p.u. and -1 p.u as the potential voltage deviation decrease outweighs any control action penalization. This regulation effort alleviates the over-voltage trend at the WTGs' POI.

WTGs' POI voltages are close with and without controller near 1 p.u. of injected active power due to the setting of active power priority. A high injection of active power barely leaves any room for injecting/absorbing reactive power. The actual injected/absorbed reactive power value can be seen in Fig. 7.

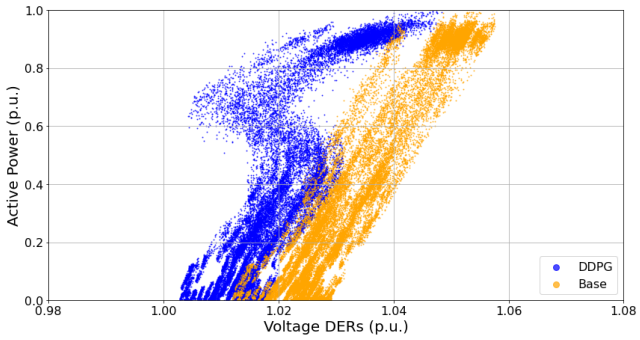

Fig. 5. WTG's active power and voltage regulation effect.

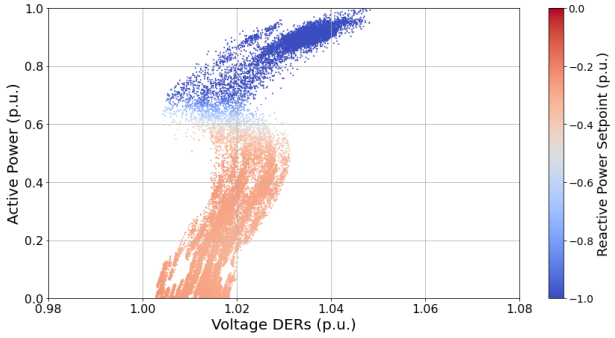

Fig. 6. WTG's reactive power setpoint for local DER voltage/active power injection.

Authorized licensed use limited to: Sandia National Laboratories. Downloaded on November 05,2025 at 19:49:16 UTC from IEEE Xplore. Restrictions apply.

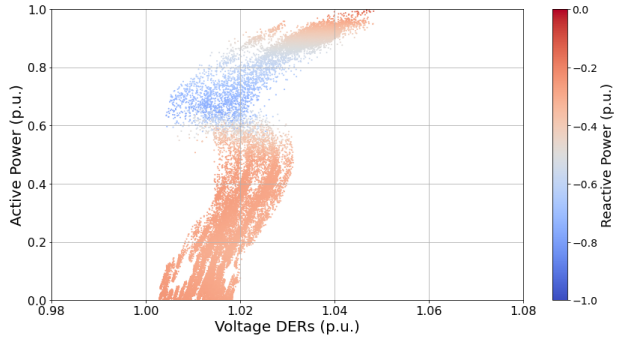

Fig. 7. WTG's measured reactive power injection for local DER voltage/active power injection.

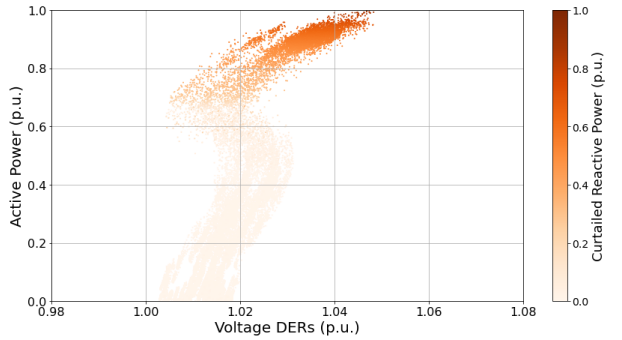

Fig. 8. WTG's reactive power curtailment for local DER voltage/active power injection.

The most noticeable voltage regulation effect occurs for an active power injection between 0.6 and 0.8 p.u., where overvoltages are significantly reduced to less than 1.02 p.u. On the contrary, the voltage regulation effect is barely noticeable for active power injections larger than 0.9 p.u. The amount of curtailed reactive power is shown in Fig. 8. For active power injections lower than 0.7 p.u., the curtailment is not noticeable.

V. Conclusions

This work proposes a Multi-Agent Reinforcement Learning (MARL) framework for cooperative voltage regulation controllers. The independent agents are located at distributed Wind Turbine Generators (WTGs) and provide voltage regulation by injecting or absorbing reactive power. The RL Agents are based on Deep Deterministic Policy Gradient (DDPG). Training and testing occurs on a Hardware-In-the-Loop setup, in which a distribution system's digital twin is executed in real-time, and measurements are sent every second to the master PC for calculating and looping back the next action. The reward is based on a weighted global voltage metric deviation and on the penalization of the control effort. The proposed controller has been proved to bring median and mean voltages closer to 1 p.u. and to reduce standard deviation. The control achieves the maximum regulation effect between 0.6 and 0.8 p.u. active power injection, as there is still enough headroom for reactive power control and voltage regulation benefits out-weight control action penalization.

References

[1] S. Report, C. Whitaker, J. Newmiller, M. E. Ropp, and B. Norris, "Distributed Photovoltaic Systems Design and Technology Requirements," 2008.

[2] R. Darbali-Zamora and M. Jimenez Aparicio, "Volt-Var Curve Control Using Wind Turbine Generators and Photovoltaic Inverters for Voltage Regulation," in *2023 IEEE Kansas Power and Energy Conference (KPEC)*, 2023.

[3] S. D. J. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi, "Multi-Agent Systems for Power Engineering Applications—Part I: Concepts, Approaches, and Technical Challenges," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1743–1752, 2007.

[4] N. Yorino, Y. Zoka, M. Watanabe, and T. Kurushima, "An Optimal Autonomous Decentralized Control Method for Voltage Control Devices by Using a Multi-Agent System," *IEEE Transactions on Power Systems*, vol. 30, no. 5, pp. 2225–2233, 2015.

[5] K. Shinya and T. Nagata, "A multi-agent systems for voltage control of distribution networks by coordination power factors of distributed generators," in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, pp. 1–5, 2016.

[6] H. Fakham, F. Colas, and X. Guillaud, "Real-time simulation of multi-agent system for decentralized voltage regulation in distribution network," in *2011 IEEE Power and Energy Society General Meeting*, pp. 1–7, 2011.

[7] Y. Pei, J. Zhao, Y. Yao, and F. Ding, "Multi-Task Reinforcement Learning for Distribution System Voltage Control With Topology Changes," *IEEE Transactions on Smart Grid*, p. 1, 2022.

[8] M. Jimenez-Aparicio and R. Darbali-Zamora, "Reactive Power Control for Fast-Acting Voltage Regulation of Distributed Wind Turbines Using Reinforcement Learning," in *2023 IEEE Kansas Power and Energy Conference (KPEC)*, 2023.

[9] D. Cao, W. Hu, J. Zhao, Q. Huang, Z. Chen, and F. Blaabjerg, "A Multi-Agent Deep Reinforcement Learning Based Voltage Regulation Using Coordinated PV Inverters," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 4120–4123, 2020.

[10] D. Hu, Y. Peng, J. Yang, Q. Deng, and T. Cai, "Deep Reinforcement Learning Based Coordinated Voltage Control in Smart Distribution Network," in *2021 International Conference on Power System Technology (POWERCON)*, pp. 1030–1034, 2021.

[11] D. Hu, Z. Ye, Y. Gao, Z. Ye, Y. Peng, and N. Yu, "Multi-Agent Deep Reinforcement Learning for Voltage Control With Coordinated Active and Reactive Power Optimization," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4873–4886, 2022.

[12] J. Wang, H. Padullaparti, F. Ding, M. Baggu, and M. Symko-Davies, "Voltage Regulation Performance Evaluation of Distributed Energy Resource Management via Advanced Hardware-in-the-Loop Simulation," *Energies*, vol. 14, no. 20, 2021.

[13] J. Berg et al., "Scaled Wind Farm Technology Facility Overview," in *32nd ASME Wind Energy Symposium*, 2014.

[14] G. E. North Piegan, R. Darbali-Zamora, and J. C. Berg, "Development and Validation of a Wind Turbine Generator Simulation Model," in *2022 North American Power Symposium (NAPS)*, pp. 1–6, 2022.

[15] J. K. Gupta, M. Egorov, and M. Kochenderfer, "Cooperative Multi-agent Control Using Deep Reinforcement Learning," in *Autonomous Agents and Multiagent Systems* (G. Sukthankar and J. A. Rodriguez-Aguilar, eds.), (Cham), pp. 66–83, Springer International Publishing, 2017.

Authorized licensed use limited to: Sandia National Laboratories. Downloaded on November 05,2025 at 19:49:16 UTC from IEEE Xplore. Restrictions apply.