Distributed Deep Deterministic Policy Gradient Agents for Real-Time Energy Management of DC Microgrid

Elutunji Buraimoh
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
eburaim@clemson.edu

Gokhan Ozkan
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
gokhano@clemson.edu

Laxman Timilsina
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
ltimils@clemson.edu

Grace Muriithi
Department of Automotive Engineering
Clemson University
Clemson, United States
gmuriit@clemson.edu

Ali Arsalan
Department of Automotive Engineering
Clemson University
Clemson, United States
aarsala@clemson.edu

Behnaz Papari
Department of Automotive Engineering
Clemson University
Clemson, United States
bpapari@clemson.edu

Ali Moghassemi
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
amoghas@clemson.edu

Christopher Edrington
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
cedring@clemson.edu

Mustafa Ozden
Department of Electrical and Computer Engineering
Clemson University
Clemson, United States
mozden@clemson.edu

2024 IEEE Sixth International Conference on DC Microgrids (ICDCM) | 979-8-3503-6363-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDCM60322.2024.10665114

***Abstract*—This paper presents a real-world military use of a microgrid with Vehicle-to-Grid (V2G) and Vehicle-to-Vehicle (V2V). This plug-and-play system delivers efficient, fast-deploying power for a contingency base within 20 minutes, generating up to 500 kW of three-phase power. It utilizes vehicle-based Internal Combustion (IC) engine generators and an Energy Storage System (ESS) to power vehicle equipment and off-board loads like shelters and communication centers. However, fluctuating off-board loads create operational challenges for this V2G-V2V microgrid. This work proposes a real-time implementation of energy management based on distributed Deep Deterministic Policy Gradient (DDPG), a deep reinforcement learning, to address this problem for continuous state and continuous action spaces. This work formulates the energy management problem as a Markov decision process, considering random load demands and fuel costs. This paper aims to minimize the microgrid's operating costs by optimizing the dispatch of onboard vehicle power generators and storage units. With simulation experiments, this work demonstrates the effectiveness of the proposed distributed deep reinforcement learning approach. The results show the distributed DDPG agents can utilize the state of the ESS and significantly decrease the microgrid's overall operation costs. This validates the practical value of distributed DDPG for economic operations in military microgrids.**

Keywords*— *Energy Management, Deep Deterministic Policy Gradient, Multi-Agent System, Microgrid, Reinforcement Learning, Vehicle-to-Grid, Vehicle-to-Vehicle.

I. Introduction

A microgrid is a localized network of interconnected power demands and distributed power sources with defined electrical borders. It functions as a single, controllable unit for operating while connected to the main grid (grid-tied) or independently (islanded). Microgrids were initially designed for crucial power needs and remote locations to enhance the reliability of the power system and expedite electrification in other sectors. There are three main microgrid configurations: AC, DC, and a hybrid combining AC and DC. These configurations are seen as viable foundations for future power system designs [1] [2]. Microgrid controllers, which manage energy using various methods, are considered the core of automated microgrid operation. A microgrid controller is a sophisticated control system, potentially made up of multiple components and sub-systems. It can sense grid conditions, monitor microgrid operations, and make adjustments to ensure continuous electricity delivery to critical loads during all microgrid operating modes (grid-tied, islanded, and transitions between them) [3]. Traditionally, microgrid energy management is demonstrated as an optimization problem focused on minimizing operational costs, such as fuel consumption and grid reliance [4]. The current research methodology can be classified into two primary categories according to the model type and solution approaches: those that rely on models and deterministic methods and those based on learning without models. Model-based deterministic approaches to microgrid energy management stand out for their precise optimization capabilities [5].

Deterministic approaches rely on detailed mathematical models of the microgrid, encompassing all components from distributed generators and energy storage systems to load profiles. This level of detail allows for accurate cost minimization, as demonstrated in [6], [7], and [8], which achieved optimal day-ahead scheduling using mixed-integer linear programming (MILP) models. [6] introduces a mathematical model that accounts for the unpredictable behavior of power consumption, renewable energy sources,

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited. OPSEC#8644 and voltage levels for optimizing energy management in unbalanced three-phase microgrids. [7] proposes a mathematical model for optimizing operations within isolated microgrids. This model depicts the microgrid as a three-phase power distribution system incorporating various power generation units, energy storage units, and renewable sources. All generation and consumption are treated as uncertain and modeled using different scenarios. [8] proposes a management algorithm that accounts for unpredictable factors to schedule the optimal use of resources within a microgrid. These resources include renewable energy sources, microturbines, fuel cells, and energy storage. The algorithm considers buying and selling energy with the main grid to minimize overall costs while maintaining power balance within the microgrid. It achieves this by formulating a mathematical objective function that aims to minimize the total operational cost. However, the strengths of model-based methods come with inherent limitations. Known system states and transfer patterns are crucial for model accuracy, posing challenges when dealing with uncertainties like fluctuating renewable energy sources and variable costs.

Learning-based approaches have emerged as a practical option in microgrid energy management, which does not require the inflexible traditional models. Unlike the deterministic approach, learning-based methods use data to directly determine the optimal control strategies directly, avoiding the need for complex system models and precise parameter estimation. This flexibility has ensured the successful application of learning-based methods across various power system applications [9] [10] [11] and [12]. This [9] introduces a novel scheduling method that utilizes forecasts to optimize energy use in real time. This method takes into account anticipated changes in the operating environment. [10] introduces a method for home energy management that uses data and a multi-agent reinforcement learning technique. This method treats the problem of scheduling home energy consumption for the next hour as a series of decisions based on past events. The goal is to achieve two objectives: minimizing the electricity bill and keeping user satisfaction high when responding to external energy demands. [12] describes developing an energy management system that manages energy in a solar panel and battery microgrid connected to the main power grid. This system considers how battery use affects its lifespan and aims to reduce strain on the battery. This scheduling aims to minimize overall costs, including buying power from the grid and battery wear and tear costs. It also maximizes the use of solar energy produced by the panels. This approach can efficiently handle uncertainties; however, the "curse of dimensionality" can pose a challenge when applied to complex microgrids with high-dimensional states and action spaces [13].

To address the limitations of existing techniques in microgrid energy management, this paper leverages the combination of deep reinforcement learning called the deep deterministic policy gradient (DDPG) algorithm. DDPG is deployed in this work for Microgrid Energy Management because it is Model-free learning that removes the need for complex system models, making it adaptable to diverse microgrid configurations. The DDPG algorithm is a powerful method within Deep Reinforcement Learning (DRL) that avoids making overly optimistic predictions. It can effectively learn the behavior of the system it's controlling and make the right adjustments even when conditions change, as explained in the following sections.

II. Methodology

The microgrid is a 600 V DC microgrid with 4 Tactical Vehicle-to-Grid Modules (TVGMs), as shown in Figure 1. Each TVGM consists of an IC engine, an Energy Storage System, and a dynamic load. Internal Combustion Engine 1, 2, and 3 are the central sources of power for the DC microgrid. IC 1 and 3 have the same generation cost because of their similar capacity (120 kW). The same is true for ICs 2 and 4 (30 kW). The DDPG is used in a multi-agent pattern to collectively achieve the goal of energy management, as shown in Figure 2 for the microgrid. Each TVGM has four onboard IC engine generators with maximum active power of 120 kW and 30 kW for TVGMs 1 and 3 and TVGMs 2 and 4, respectively. The power output for the onboard generators is given as $P_{n,t}{}^{IC,}$ and limits are given in Equation 1. The fuel cost of the IC engines is expressed as a function of the power output as given in Equation 2, where the coefficients alpha, beta, and gamma are IC engine generator characteristics, which are the same for 120 kW generators 1 and 3 and 30 kW generators 2 and 4.

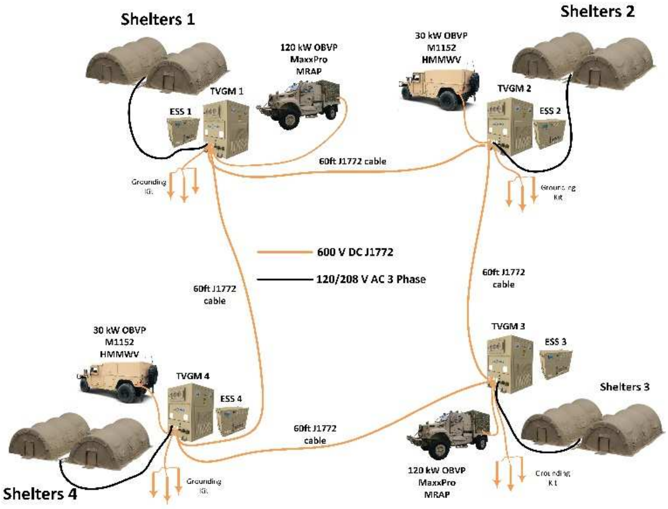

Fig. 1. Military Based V2G/V2V Microgrid.

The real-time energy management is formulated as a Markov Decision Process (MDP) problem with unknown transition probability and uncertainty in the dynamic load.

$$P_{n,min}^{IC} < P_{n,t}^{IC} < P_{n,max}^{IC} \tag{1}$$

$$C_{n,t}^{IC} = \left[\alpha_n\left(P_{n,t}^{IC}\right)^2 + \beta_n\left(P_{n,t}^{IC}\right) + \gamma_n\right]\cdot\Delta t \tag{2}$$

Each TVGM has an equal capacity Energy Storage System (ESS) of 22.8 kWh. The output of the ESS is represented as Pn,t ESSand the value falls within the limit expressed in Equation 3, which infers charging or discharging of the ESS within its maximum power for charging or discharging. Similarly, the state of charge of the ESS is kept within the constraint of the ESS to function, as shown in Equation 4.

$$-P_{n,max}^{ESS} \leq P_{n,t}^{ESS} \leq P_{n,max}^{ESS} \tag{3}$$

$$SoC_{n,min} \leq SoC_{n,t} \leq SoC_{n,max} \tag{4}$$

The constraint of the power balance in the microgrid is given by Equation 5, in which the total power demand within the microgrid must match the total power demanded.

$$\sum P_{n,t}^{IC} = \sum P_{n,t}^{L} + \sum P_{n,t}^{ESS} \tag{5}$$

The objective of this real-time energy management is to minimize the cost of fuel consumption for the onboard generators of the microgrid, and thus, the state of the system is given by Equation 6. Thus, the system's action is given by Equation 7, which represents the power references for the IC engine onboard generators and the ESS power reference to charge or discharge.

$$s_t = \left(P_{n,t}^{L}, SoC_{n,t}, t\right),\ s_t \epsilon\ S \tag{6}$$

$$a_t = \left(P_{n,t}^{IC}, P_{n,t}^{ESS}\right),\ a_t \epsilon\ A(s_t) \tag{7}$$

The reward is expressed as the negative of the fuel cost function of the IC engines, as well as the penalty cost associated with the use of the ESS, as shown in Equation 8. The MDP objective in Equation 9 is to minimize the fuel cost by determining the best sequence of actions that deliver the maximum cumulative reward expressed as π^*

$$r_t(s_t, a_t) = -\left[\sum C_{n,t}^{IC} + \sum P_{n,t}^{ESS}\right] \tag{8}$$

$$V^{\pi*}(s_o) = \max \mathbb{E}_{\pi}\left[\sum \gamma^{t} \cdot r_{t+1} | s_o\right] \tag{9}$$

Each DDPG provides continuous action space, enabling precise energy scheduling without discretization limitations. DDPG, as a policy gradient approach, learns optimal dispatch policies directly from interactions with the environment, continuously improving its performance. DDPG uses iteration techniques to efficiently utilize sampled data efficiently, promoting faster convergence and stable learning. In this work, the DDPG is adapted for V2G-V2V microgrid optimization. Firstly, for the state Representation, the Microgrid state is captured by features like off-board load demand and ESS state of charge. This design accommodates real-world uncertainties without relying on specific uncertainty distributions. Secondly, energy management is designed to minimize operational costs, guiding each DDPG policy toward economically optimal decisions for the reward function.

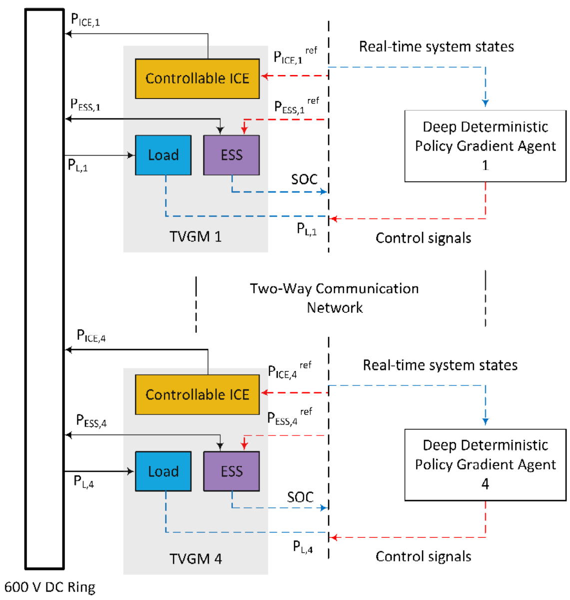

Fig. 2. Distributed DDPG Agents in Military Based V2G/V2V Microgrid.

In the case of reinforcement learning, the problem of optimization expressed in Equation 9 is re-expressed as shown in Equation 10 with $V^{\pi*}$as the required optimal cumulative reward and $Q^{\pi*}$ as the function of action-value. This function as shown in Equation 11 is a Bellman Equation that is solved to derive the optimal policy π^* as shown in Equation 12.

$$V^{\pi*}(s_o) = \max Q^{\pi*}(s_t, a_t) \tag{10}$$

$$Q^{\pi*}(s_t, a_t) = \mathbb{E}_{\pi}[r_t + \gamma \cdot \max Q^{\pi*}(s_{t+1}, a_{t+1})] \tag{11}$$

$$\pi^*(s_t) = argmax\ Q^{\pi*}(s_t, a_t) \tag{12}$$

A custom Deep Neural Network (DNN) architecture is employed to handle the specific features and action space of the microgrid energy management problem. This approach utilizes a DNN to approximate the optimal action-value function, denoted by $Q^{\pi*}$. The DNN is trained iteratively to refine its estimate over time. Equation 13 defines the update rule for the critic's parameters based on a specific loss function with B as the batch size and y as the target value.

$$L(\theta^Q) = \frac{1}{B}\sum\left(y_i - Q(s, a|\theta^Q)\right)^2 \tag{13}$$

The DDPG algorithm leverages an actor-critic architecture. This architecture comprises two key elements: the actor and the critic. The critic, denoted by *Q(s, a)*, is trained using the Bellman equation, similar to Q-learning. The DDPG algorithm trains the DNN policy network iteratively using real-time data and the defined reward function

For implementation and validation, the trained DDPG policy is used in real-time, taking the current microgrid state as input and providing continuous scheduling decisions as output (which in turn serves as the power reference for the onboard IC engines). The adaptability of the DNN automatically adjusts to different system dimensions, making it scalable for complex, high-dimensional scenarios. Extensive simulations employing real power system data will validate the efficiency of the proposed DDPG-based approach in achieving cost-optimized energy management for the V2G-V2V microgrid.

III. Results and Discussion

Figure 3 depicts the proposed method's total reward convergence behavior throughout the training process. These curves demonstrate the method's capability to learn and enhance the reward function. This implies the DDPG approach achieves a stable policy within the simulated environment. As shown in Figure 3, an episode refers to a single complete interaction between the agent and the environment in DDPG training.

The episode reward is the total reward the agent accumulates throughout that episode, representing the agent's overall performance in a single run. The average reward is the average of the episode rewards calculated over a specific number of episodes. It provides a broader picture of the agent's performance over time and helps smooth out fluctuations in individual episode rewards. Tracking the average reward is crucial to monitor the agent's learning progress, and as shown in Figure 3, the trend shows that the agent learns better policies. Episode Q0 refers to the estimated discounted long-term reward for the initial state of an episode, as predicted by the critic network in DDPG. The critic network aims to learn the value (reward) of being in a particular state and taking a specific action. Episode Q0 represents the critic's estimate of the total discounted reward the agent can expect to receive throughout the entire episode, starting from the initial state. As shown in Figure 3, the trend indicates that the agent learns better policies, reflecting its understanding of maximizing long-term rewards.

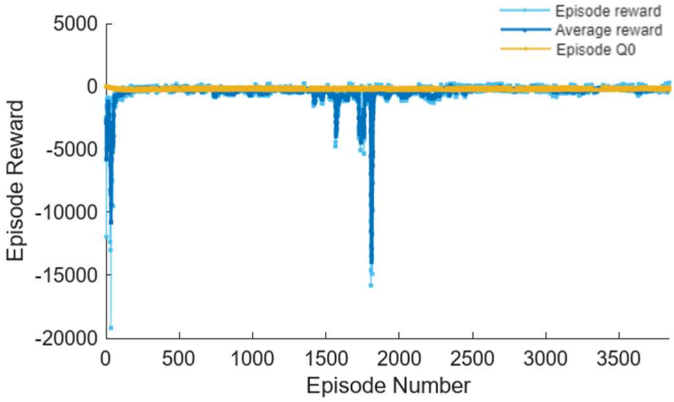

Fig. 3. Cumulative Reward Function.

Figures 4, 5, 6, and 7 show the power dispatch across the four TVGMs of the microgrid. The distributed DDPG agents work together to achieve power balance within the microgrid while exchanging power among themselves.

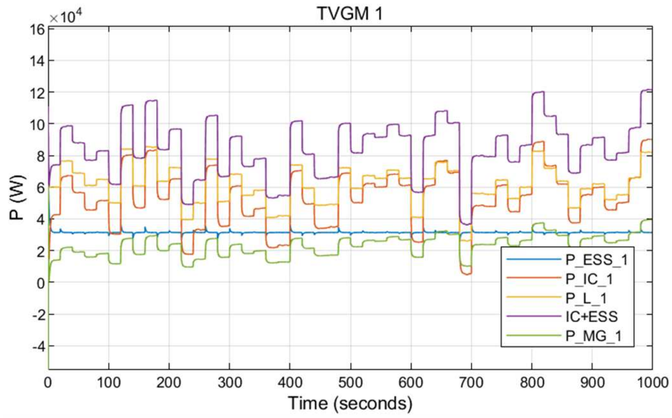

Fig. 4. Power Dispatch in TVGM 1.

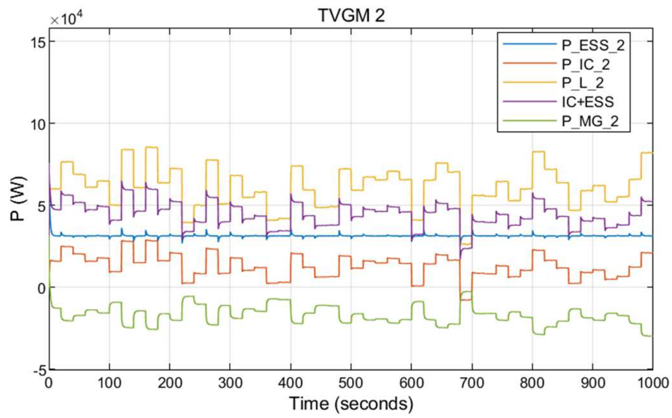

Fig. 5. Power Dispatch in TVGM 2.

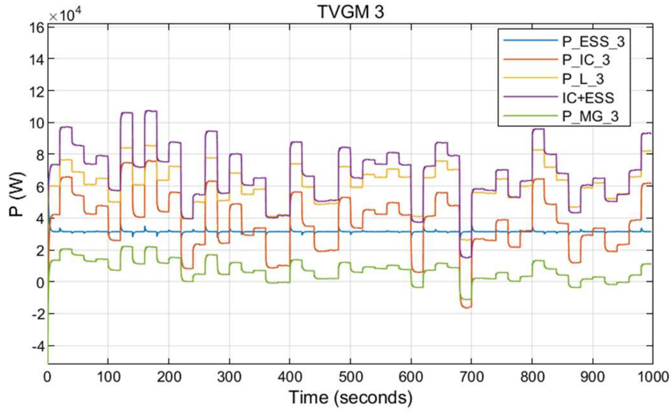

Fig. 6. Power Dispatch in TVGM 3.

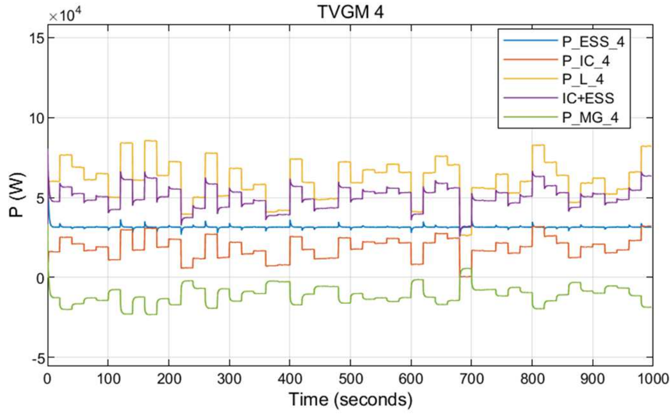

Fig. 7. Power Dispatch in TVGM 4.

The proposed approach can increase the reward when the load is dynamic and unpredictable. This shows that the proposed Distributed DDPG approach learns a stable policy and solves the energy management problem. The results show that the proposed DDPG method works well under dynamic load conditions. The Energy Storage Systems 1, 2, 3, and 4 are discharged and charged, corresponding to the power they can store and the cost function of the four IC engines. Based on the fuel cost and the charging process of the Energy Storage System, the Energy Storage System can respond to changes in fuel cost. The Energy Storage Systems 1, 2, 3, and 4 are charged at a low cost of fuel, and then their State of Charge is increased. They are discharged at a high fuel cost, and then their State of Charge decreases. With the Energy Storage System, the entire Vehicle-to-Vehicle/Vehicle-to-Grid Microgrid can also respond to changes in fuel cost. During low fuel costs, if the internal combustion engine and energy storage system of a particular TVGM cannot meet the local TVGM load demand, power is exchanged between the TVGMs. Consequently, power balance is achieved across the entire microgrid, as shown in Figure 8. The simulation result is shown in Figure 8 to demonstrate that the distributed DDPG can learn and achieve an economic dispatch.

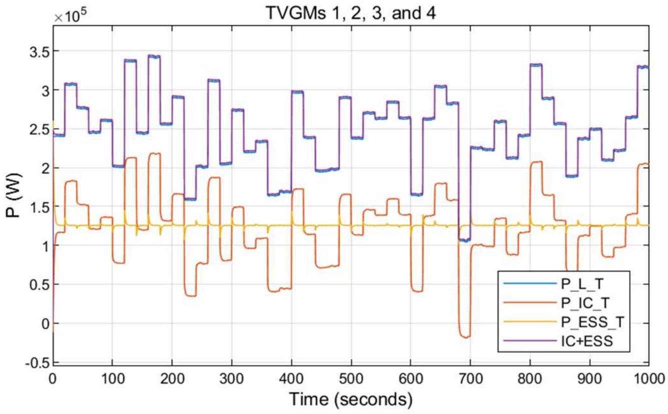

Fig. 8. Power Dispatch across the entire TVGMs in the Microgrid.

IV. Conclusion

Military microgrids, which utilize vehicle-to-grid (V2G) and vehicle-to-vehicle (V2V) technology for power distribution, require intelligent energy management strategies. This paper addresses this challenge by proposing a novel approach based on distributed deep reinforcement learning. This method offers significant advantages over traditional techniques: Reduced Complexity: Unlike conventional methods that rely on intricate system models, this approach eliminates the need for such models, making it adaptable to diverse military microgrid configurations; and Real-Time Optimization: The proposed method operates in real-time, dynamically adjusting power generation and storage usage based on fluctuating off-board loads. This ensures efficient energy utilization within the microgrid. This work introduces a real-time energy management system for V2G/V2V DC microgrids employed in military applications. The system's core leverages a distributed Deep Deterministic Policy Gradient (DDPG) algorithm, a deep reinforcement learning technique suited for continuous state and action spaces. The system considers factors like random load demands and fuel costs by formulating the energy management problem as a Markov Decision Process (MDP). The primary objective is to minimize the microgrid's operational expenses by optimizing power generation dispatch from onboard vehicle internal combustion (IC) engines and energy storage units (ESS). The effectiveness of the proposed distributed deep reinforcement learning approach is validated through extensive simulations. The results demonstrate how the DDPG agents, equipped with real-time information on the ESS state, can significantly reduce the overall operating costs of the microgrid. This reinforces the practical value of distributed DDPG for achieving economic operation in military microgrids.

Acknowledgment

This work is supported by Clemson University's Virtual Prototyping of Autonomy Enabled Ground Systems (VIPR-GS), under Cooperative Agreement W56HZV-21-2-0001 with the US Army DEVCOM Ground Vehicle Systems Center (GVSC).

References

[1] X. Liu, T. Zhao, H. Deng, P. Wang, J. Liu, and F. Blaabjerg, "Microgrid Energy Management with Energy Storage Systems: A Review," *CSEE J. Power Energy Syst.*, vol. 9, no. 2, pp. 483–504, 2023, doi: 10.17775/CSEEJPES.2022.04290.

[2] E. Buraimoh, I. E. Davidson, and F. Martinez-Rodrigo, "Fault ride-through enhancement of grid supporting inverter-based microgrid using delayed signal cancellation algorithm secondary control," *Energies*, vol. 12, no. 20, pp. 7–8, 2019, doi: 10.3390/en12203994.

[3] E. Buraimoh, G. Ozkan, L. Timilsina, P. K. Chamarthi, B. Papari, and C. S. Edrington, "Overview of Interface Algorithms, Interface Signals, Communication and Delay in Real-Time Co-Simulation of Distributed Power Systems," *IEEE Access*, vol. 11, no. July, pp. 103925–103955, 2023, doi: 10.1109/ACCESS.2023.3317250.

[4] L. Timilsina *et al.*, "A Real-Time Prognostic-Based Control Framework for Hybrid Electric Vehicles," *IEEE Access*, vol. 11, no. November, pp. 127589–127607, 2023, doi: 10.1109/ACCESS.2023.3332689.

[5] A. Arsalan *et al.*, "A Resilient Distributed Control and Energy Management Approach for DERs and EVs with Application to EV Charging Stations," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 2115–2120, 2023, doi: 10.1016/j.ifacol.2023.10.1114.

[6] J. A. A. Silva, J. C. López, N. B. Arias, M. J. Rider, and L. C. P. da Silva, "An optimal stochastic energy management system for resilient microgrids," *Appl. Energy*, vol. 300, no. April, p. 117435, 2021, doi: 10.1016/j.apenergy.2021.117435.

[7] P. P. Vergara, J. C. López, M. J. Rider, H. R. Shaker, L. C. P. da Silva, and B. N. Jørgensen, "A stochastic programming model for the optimal operation of unbalanced three-phase islanded microgrids," *Int. J. Electr. Power Energy Syst.*, vol. 115, no. June 2019, p. 105446, 2020, doi: 10.1016/j.ijepes.2019.105446.

[8] N. Eghbali, S. M. Hakimi, A. Hasankhani, G. Derakhshan, and B. Abdi, "Stochastic energy management for a renewable energy based microgrid considering battery, hydrogen storage, and demand response," *Sustain. Energy, Grids Networks*, vol. 30, p. 100652, 2022, doi: 10.1016/j.segan.2022.100652.

[9] M. Ren, X. Liu, Z. Yang, J. Zhang, Y. Guo, and Y. Jia, "A novel forecasting based scheduling method for household energy management system based on deep reinforcement learning," *Sustain. Cities Soc.*, vol. 76, no. June 2021, p. 103207, 2022, doi: 10.1016/j.scs.2021.103207.

[10] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A Multi-Agent Reinforcement Learning-Based Data-Driven Method for Home Energy Management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, 2020, doi: 10.1109/TSG.2020.2971427.

[11] A. H. Yakout, H. M. Hasanien, R. A. Turky, and A. E. B. Abu-Elanien, "Improved reinforcement learning strategy of energy storage units for frequency control of hybrid power systems," *J. Energy Storage*, vol. 72, no. PA, p. 108248, 2023, doi: 10.1016/j.est.2023.108248.

[12] G. Muriithi and S. Chowdhury, "Optimal energy management of a grid-tied solar pv-battery microgrid: A reinforcement learning approach," *Energies*, vol. 14, no. 9, 2021, doi: 10.3390/en14092700.

[13] G. Muriithi and S. Chowdhury, "Deep Q-network application for optimal energy management in a grid-tied solar PV-Battery microgrid," *J. Eng.*, vol. 2022, no. 4, pp. 422–441, 2022, doi: 10.1049/tje2.12128.