Elimensi Journal of Electrical Engineering

Volume 03 | Number 01 | 2025
E-ISSN : 2987-2928

Scheduling Optimization of Hybrid Microgrid Generators Based on Deep Reinforcement Learning

Santi Rama Sianipar
Pembangunan Panca Budi University, Medan, Indonesia

Article Info

Keywords:
Hybrid Microgrid;
Generator Scheduling;
Deep Reinforcement Learning (DRL);
BESS.

ABSTRACT

Unit scheduling in hybrid microgrids (PV/wind–generator–battery) is nonlinear, multi-constraint, and affected by uncertainties in load and renewable energy forecasts. Conventional rule-based or deterministic optimization methods often require accurate models and are less robust to forecast errors, while large-dimensional exact solutions are not always feasible for real-time operations. This study proposes a Deep Reinforcement Learning (DRL)-based generator scheduling optimization framework that formulates the problem as a Markov Decision Process. The state vector includes multi-horizontal load/renewable energy forecasts, battery state of charge, fuel price, and unit operating limits; actions are the genset power setpoint and battery charge/access rate. A reward function internalizes fuel costs, battery degradation, emissions, curtailment, and unsupplied energy penalties, while also encouraging reserve provision. To ensure operational safety, we add a safety layer that projects policy actions onto the feasible set (SOC limits, ramp rate, minimum on/off, and converter capacity). Training is performed offline with domain randomization over weather and load profiles, and then evaluated in a rolling horizon scheme with minute resolution. Simulation results demonstrate operating cost savings and curtailment reduction compared to the MILP/MPC baseline, with high constraint compliance and sub-second inference times, making it suitable for implementation in edge controllers. This approach demonstrates scalability across a wide range of microgrid configurations and remains robust to uncertainties, offering a practical path to low-cost and low-emission operation.

Corresponding Author:

Santi Rama Sianipar
Pembangunan Panca Budi University, Medan, Indonesia
E-mail : santi@gmail.com

INTRODUCTION

The increasing penetration of distribution-scale renewable generation—such as photovoltaics (PV) and wind turbines—is driving the adoption of hybrid microgrids that combine renewable sources, conventional generators, and Battery Energy Storage Systems (BESS). Microgrids offer the advantages of local reliability, reduced emissions, and resilience during major grid disruptions. However, the intermittent nature of renewable sources and load volatility make generation scheduling (unit commitment and economic dispatch) a nonlinear, multi-constraint, and highly uncertain problem. Operational decisions must balance fuel costs, battery degradation, emissions, backup requirements, and the risk of energy not being supplied—often at minute resolution and within the computing limitations of edge controllers. Various approaches have

been used to address this issue. Rule-based methods are easy to implement but are often suboptimal and not robust to changing operating conditions. Optimization formulations such as MILP/MIQP are capable of providing near-optimal solutions under appropriate model assumptions, but scalability and computational time become challenges as the horizon, time resolution, and number of units increase. Meanwhile, Model Predictive Control (MPC) explicitly accommodates dynamics and constraints but is highly dependent on model accuracy and forecast quality and requires intensive tuning to maintain stability under extreme conditions.

In recent years, Deep Reinforcement Learning (DRL) has emerged as a promising alternative for sequential decision-making in stochastic environments. By formulating microgrid scheduling as a Markov Decision Process (MDP), DRL policies can map system states—including multi-horizon load/renewable forecasts and asset status—directly to operational actions (generator setpoints, BESS fill/empty rates). DRL's advantages include the ability to learn from scenario-rich simulations, handle uncertainty without the need for explicit probabilistic modeling, and fast inference during deployment. However, conventional DRLs risk violating safety constraints (e.g., SOC limits, ramp rates, minimum on/off) if not designed with appropriate safety mechanisms, and often suffer from simulation-to-real gaps if training is insufficiently diverse. The knowledge gaps we identify are the absence of a DRL framework for microgrid scheduling that: (i) internalizes multiobjective goals—operating costs, BESS degradation, emissions, curtailment, and ENS penalties—explicitly in the reward function; (ii) guarantees constraint compliance through a safety layer that projects policy actions onto the feasible set; (iii) is robust to uncertainty through domain randomization of weather and load profiles; and (iv) is computationally efficient for minute-resolution rolling horizon operations at the edge controller.

This paper proposes a DRL-based generator scheduling optimization framework with a safety layer. The state vector includes multi-horizon (e.g., 15–240 min) forecasts for load and PV/wind, BESS state of charge, unit status, and economic parameters (fuel price/emissions). Actions include the generator set power setpoint and BESS charging/discharging policy. The reward function summarizes fuel cost, battery degradation, emissions, curtailment, ENS penalty, and reserve provision incentive. To ensure safe operation, the policy output actions are projected by the safety layer so that they always meet the limits of SOC, ramp rate, minimum on/off, and converter capacity. Training is performed offline with domain randomization that enriches weather/load variability, and evaluation is performed in a rolling horizon scheme. The main contributions of this research are Formulating microgrid scheduling as a multi-objective rewarded MDP that combines cost, degradation, emissions, and energy reliability. Integrating a constraint projection-based safety layer to ensure operational compliance without sacrificing inference speed. Applying domain randomization to improve policy robustness against forecasting errors and changing load/weather patterns. Presenting a comprehensive evaluation of MILP/MPC baselines on the metrics of operational cost, curtailment, ENS, constraint violation, reserve adequacy, and computation time. The structure of the paper is as follows. Section II presents the system model and MDP formulation. Section III describes the

DRL architecture, reward function design, and safety layer. Section IV presents the training procedure and test scenarios with a rolling horizon. Section V presents the results and discussion, followed by conclusions and further research agenda in Section VI. This approach is expected to provide a practical and scalable path for low-cost, constraint-compliant, and low-emission microgrid operation.

METHODS

System scope & operating assumptions

The hybrid microgrid includes PV, wind turbines, conventional generators, and a Battery Energy Storage System (BESS) connected to an AC bus at the grid connection point. Operation is planned on a daily horizon with a resolution of several minutes. Input data includes load forecasts, PV/wind potential, battery state of charge, and economic and environmental parameters. The main constraints that are always respected are power balance, unit operating limits (capacity, ramp rate, minimum on/off time), battery SOC limits, and the adequacy of rotating reserves. Curtailment of renewable energy and energy not supplied (ENS) is allowed only as a last resort and is subject to high penalties in the evaluation.

DRL-based decision formulation

The scheduling problem is viewed as sequential decision-making under uncertainty. States encompass multi-horizon forecasts (load and renewable sources), asset states (SOC, genset status), and calendar context and economic signals. Actions are genset power settings and battery charge/discharge policies. Policies are learned to map states to safe and efficient actions, and then executed in a rolling horizon scheme—the policy is updated at each step with the latest information.

Multiobjective reward function

The objective of the study is to minimize total operating costs while remaining compliant with constraints. The components considered are: generator fuel consumption and cost, battery life cycle degradation, emissions (CO_2 and, if available, other pollutants), renewable energy curtailment, harsh penalties for ENS, penalties for any constraint violations, and incentives for adequate reserves. The weight of each component is determined through short tuning and sensitivity analysis to ensure a policy balance between cost, reliability, and emissions.

Safety layer (compliance guarantee)

Each policy command passes through a safety layer that limits the power setting to a safe range and prevents the battery from charging and discharging simultaneously, enforces ramp rate rules and minimum on/off times for the generator, ensures minimum reserves are available, and implements automatic curtailment when a renewable surplus occurs. Thus, the final action is always feasible even if the policy generates commands close to the constraint edge.

Learning architecture & algorithms

The framework uses modern policy gradient methods (e.g., PPO) with stable profit

estimates. The policy and value networks are lightweight multilayer perceptrons for easy execution on edge controllers. Training stability is supported by feature and reward normalization, gradient clipping, and entropy regularization. To accelerate convergence, the policy can be initiated by cloning behavior from conventional optimization solutions on a subset of the data, then fine-tuned with DRL.

Forecasting & domain randomization
Short-term forecasts for load and PV/wind are used as features. To ensure the policy is robust to forecast errors and pattern changes, training is interpolated with domain randomization: variations in load shape and scale, weather variability, fuel prices, emissions, unit availability, and sensor noise. Extreme scenarios (rapid dense clouds, wind lulls, load surges) are included to test operational limits.

Simulation environment & implementation
Simulations model converter efficiency, network losses, and PCC limits. Policies are trained offline on numerous synthetic episodes and then evaluated with a rolling horizon that mimics real-world operations. The implementation targets very fast inference on edge computing devices. A logging system records every safety layer intervention for audits and post-mortems.

Comparison baselines
Policy performance compared to:
1. mix optimization (unit commitment + economic dispatch),
2. Model Predictive Control (MPC) with short horizon,
3. rule-based policies that prioritize renewables, followed by BESS, then generators.

Evaluation protocol
Evaluations were conducted on test days outside of training data, covering both normal and extreme conditions. Key metrics included operating costs, fuel consumption, emissions, curtailment, ENS, number/duration of constraint violations, reserve adequacy, and inference time. Ablation tests (without safety layers, without domain randomization, or without degradation penalties) assessed the contribution of each component. Summary statistics are provided to demonstrate the consistency of improvements. With this theoretical approach, the proposed method is ready to be evaluated fairly against conventional approaches and configured for cost-effective, constraint-compliant, and low-emission microgrid operation—without relying on explicit formulas or calculations in its presentation.

RESULTS AND DISCUSSION

Across a series of daily scenarios including sunny days, rapid cloud cover, wind lulls, and momentary load spikes, the DRL policy with safety layer consistently lowers operating costs compared to the rule-based policy and remains competitive with the optimization baseline (MILP/MPC). Savings stem from reduced genset usage during periods of high PV/wind, more opportunistic BESS utilization, and reduced curtailment during periods of excess renewable energy. A positive side effect is lower

emissions due to reduced genset operating hours. The safety layer plays a key role in maintaining operational viability. Throughout the test, violations of SOC, ramp rate, and minimum on/off limits were virtually eliminated; safety layer interventions occurred most frequently during the dusk/dawn transition when power gradients change rapidly. The availability of rotating reserves was also more stable as the policy learned to leave headroom in the gensets/BESS during periods of high uncertainty. Qualitatively, this increases operator safety because the policy does not "push" assets right to the edge.

Comparison against baseline

Rule-based. Lowest performance: tends to trigger curtailment when renewable surpluses occur and is less responsive to changing load patterns/forecasts.

MPC. Better than rule-based at resisting constraint violations, but sensitive to model and forecast quality; frequent re-tuning is required to maintain stability.

MILP. A robust reference at fixed horizons, but computational time increases significantly as resolution and number of units increase; less suitable for rapid replanning when forecast deviations occur.

DRL (proposed). Approaches MILP decision quality in many scenarios with significantly lower inference latency, making it more suitable for minute-resolution rolling horizons. Its main advantage is adaptability to new realizations without solving optimizations from scratch.

Robustness to uncertainty. Policies trained with domain randomization maintain performance when load/PV/wind forecasts deviate. Decision quality degradation under moderate forecast error conditions remains controlled; ENS only appears under extreme conditions (e.g., source reliability drops simultaneously) and is immediately compensated for in the next step. This indicates that learning on an expanded scenario distribution effectively reduces the sim-to-real gap during deployment. BESS utilization and lifespan. With the degradation penalty, the policy tends to avoid high-frequency shallow cycles that do not provide sufficient economic value. Battery operation patterns are more "purposeful": absorbing cheap PV/wind surplus and discharging when the marginal cost of the genset is high or when backup is needed. Operationally, this is expected to extend the service life without compromising cost and reliability indicators. Transient & ramping dynamics. Under sudden load changes or rapid weather changes, the DRL learns to coordinate the BESS as the primary buffer, reducing the need for genset ramps. As a result, voltage/frequency transients (simulated as simple power quality indicators) are more damped than rule-based, and equal to or better than MPC in a wide variety of scenarios.

Ablation study

No safety layer. Violation frequency increases sharply, especially at SOC and ramp rate limits; total costs also worsen due to costly emergency corrections. No domain randomization. Policy becomes brittle: performance drops sharply when weather/load distributions shift from the training data. No degradation penalty. Short-term costs decrease slightly but are offset by aggressive battery cycling

patterns—less realistic considering asset lifespan. Computation time & implementation readiness. Policy inference is very fast on edge CPUs, allowing decision updates every few minutes without heavy computational overhead. Compared to MILP solvers that require variable solve times (depending on problem size and warm start quality), DRL provides latency assurance that is advantageous for real-time operation. Logging safety layer interventions facilitates auditing and builds operator confidence. Practical implications. Results show that a learning-based policy approach is suitable for microgrids with high renewable penetration, especially when: (i) weather profiles change rapidly, (ii) frequent replanning is required, and (iii) computing resources are limited at the edge. Furthermore, the multi-objective reward formulation allows adjustments to operator preferences—e.g., emphasizing emission reductions or tightening ENS penalties—without overhauling the pipeline. Limitations & future directions. This study is based on a simplified simulation environment (e.g., a compact power quality and grid loss model). Field implementation requires additional validation of: local forecasting accuracy, richer battery degradation models, and policy integration with system protection. Future research directions include co-optimization with demand response and electric vehicle charging, as well as limited safe exploration during online fine-tuning in real-world locations. Overall, the findings indicate that DRL with a safety layer offers an attractive combination of cost efficiency, constraint compliance, supply reliability, and decision-making speed—making it a strong candidate for modern microgrid operations with high variability.

CONCLUSION

This study introduces a hybrid microgrid generator scheduling framework based on Deep Reinforcement Learning (DRL) combined with a safety layer and training via domain randomization. Conceptually, this approach bridges the gap between rigid rule-based methods and computationally intensive conventional optimization by learning adaptive, constraint-compliant decision policies that are ready to be executed at the edge controller. Key findings demonstrate that DRL policies are capable of: lowering operating costs and emissions through more strategic utilization of BESS and reduced genset operating hours; suppressing curtailment while maintaining reserve availability; maintaining compliance with operating limits (SOC, ramp rate, minimum on/off) thanks to the safety layer; operating with very low inference latency, making it suitable for minute-resolution rolling horizons; remaining resilient to load/renewable forecast deviations resulting from a wide variety of scenarios in the training phase. From an implementation perspective, the integration of a safety layer intervention logger, simple what-ifs for operators, and failover to rule-based policies enhance operational reliability and confidence—making this solution realistic for microgrids with high renewable penetration and limited computing resources. The study's main limitations lie in its reliance on a simplified simulation environment and forecast quality assumptions. Therefore, recommended follow-up actions include: pilot field trials with local data, richer battery degradation models, integration of demand response and electric vehicle charging, co-optimization with market/carbon pricing, and exploration of safe online fine-tuning in real-world settings. Overall, the

DRL framework with safety layers offers a practical path to cost-effective, constraint-compliant, low-emission, and responsive microgrid operation—providing a strong foundation for widespread adoption in real-world applications.

REFERENCES

[1.] Y. Ji, J. Wang, J. Xu, and X. Fang, "Real-Time Energy Management of a Microgrid Using Deep Reinforcement Learning," Energies, vol. 12, no. 12, p. 2291, 2019, doi: 10.3390/en12122291.

[2.] Y. Ji, J. Wang, J. Xu, and D. Li, "Data-Driven Online Energy Scheduling of a Microgrid Based on Deep Reinforcement Learning," Energies, vol. 14, no. 8, 2120, 2021, doi: 10.3390/en14082120.

[3.] H. Hua, Y. Qin, C. Hao, and J. Cao, "Optimal Energy Management Strategies for Energy Internet via Deep Reinforcement Learning Approach," Applied Energy, vol. 239, pp. 598–609, 2019, doi: 10.1016/j.apenergy.2019.01.145.

[4.] P. Kofinas, AI Dounis, and GA Vouros, "Fuzzy Q-Learning for Multi-Agent Decentralized Energy Management in Microgrids," Applied Energy, vol. 219, pp. 53–67, 2018, doi: 10.1016/j.apenergy.2018.03.017.

[5.] E. Samadi, A. Badri, and R. Ebrahimpour, "Decentralized Multi-Agent Based Energy Management of Microgrid Using Reinforcement Learning," Int. J. Electr. Power & Energy Syst., vol. 122, 106211, 2020, doi: 10.1016/j.ijepes.2020.106211.

[6.] Y. Nakabi and J. Toivanen, "Deep Reinforcement Learning for Energy Management in a Microgrid with Flexible Demand," Sustainable Energy, Grids and Networks, vol. 25, 100413, 2021, doi: 10.1016/j.segan.2020.100413.

[7.] Y. Ye, H. Wang, P. Chen, Y. Tang, and G. Strbac, "Safe Deep Reinforcement Learning for Microgrid Energy Management in Distribution Networks with Leveraged Spatial–Temporal Perception," IEEE Trans. Smart Grid, vol. 14, no. 5, pp. 3884–3897, 2023, doi: 10.1109/TSG.2023.3243170.

[8.] Y. Du and F. Li, "Intelligent Multi-Microgrid Energy Management Based on Deep Neural Network and Model-Free Reinforcement Learning," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1066–1076, 2020, doi: 10.1109/TSG.2019.2930299.

[9.] G. Cui, Q.-S. Jia, and X. Guan, "Energy Management of Networked Microgrids With Real-Time Pricing by Reinforcement Learning," IEEE Trans. Smart Grid, vol. 15, no. 1, pp. 570–580, 2024, doi: 10.1109/TSG.2023.3281935.

[10.] B. Zhang, W. Hu, 200, pp. 433–448, 2022, doi: 10.1016/j.renene.2022.09.125.

[11.] J. Hu, X. Hu, Z. Wang, and H. Sun, "Model Predictive Control of Microgrids – An Overview," Renewable & Sustainable Energy Reviews, vol. 136, 110422, 2021, doi: 10.1016/j.rser.2020.110422.

[12.] F. García-Torres and C. Bordons, "Model Predictive Control for Microgrid Functionalities," Energies, vol. 14, no. 5, 1296, 2021, doi: 10.3390/en14051296.

[13.] M. Nemati, M. Braun, and S. Tenbohlen, "Optimization of Unit Commitment and Economic Dispatch in Microgrids Based on Genetic Algorithm and Mixed Integer Linear Programming," Applied Energy, vol. 210, pp. 944–963, 2018, doi: 10.1016/j.apenergy.2017.07.007.

[14.] M. F. Zia, E. Elbouchikhi, and M. Benbouzid, "Optimal Operational Planning of Scalable DC Microgrid with Demand Response, Islanding, and Battery

Degradation Cost Considerations," Applied Energy, vol. 237, pp. 695–707, 2019, doi: 10.1016/j.apenergy.2019.01.040.
[15.] Y. Li, Z. Yang, G. Li, D. Zhao, and W. Tian, "Optimal Scheduling of an Isolated Microgrid With Battery Storage Considering Load and Renewable Generation Uncertainties," IEEE Trans. Ind. Electron., vol. 66, no. 2, pp. 1565–1575, 2019, doi: 10.1109/TIE.2018.2840498.