



POLITECNICO
MILANO 1863



Leveraging Reinforcement Learning for Microgrid Optimal Control

D. Salaorni, F. Bianchi, F. Trovò, M. Restelli



INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS

IJCNN2025

30 JUNE - 5 JULY 2025 | ROME, ITALY

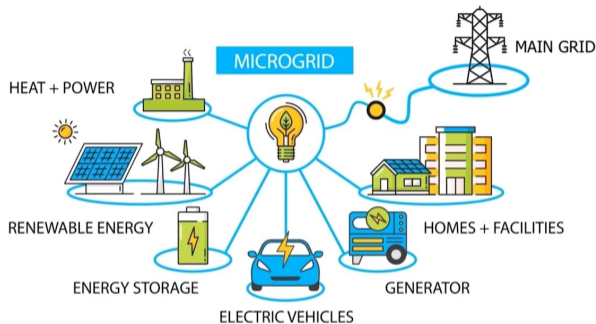


INTERNATIONAL NEURAL NETWORK SOCIETY

International Joint Conference on Neural Network (IJCNN 2025)

Motivation

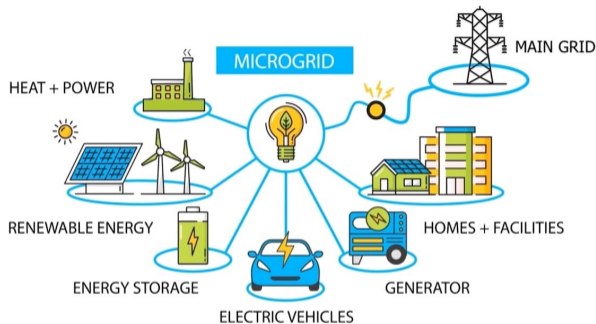
The Microgrid System



Optimize energy management in microgrids means

Motivation

The Microgrid System

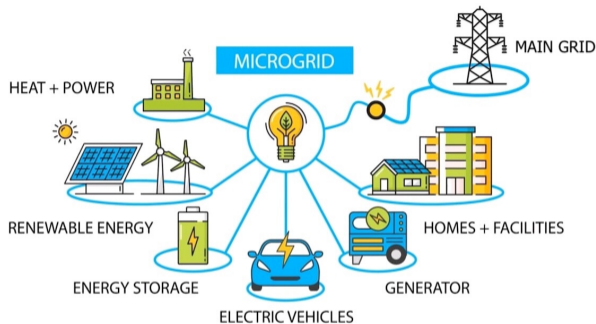


Optimize energy management in microgrids means

- Dealing with the **uncertainty** in energy generation and consumption

Motivation

The Microgrid System



Optimize energy management in microgrids means

- Dealing with the **uncertainty** in energy generation and consumption
- **Minimizing energy-related costs**, which depend on:
 - the **spread** between buying and selling energy prices
 - the **degradation** of the battery and its replacement cost

The Proposed Solutions

From the Literature

- Energy management with **no market prices** (Liu et al. [2021])
- Limited battery simulation with **no degradation costs** (Shojaeighadikolaei et al. [2021], Kolodziejczyk et al. [2021], Domínguez-Barbero et al. [2020], Chen and Bu [2019], Guo et al. [2022])

The Proposed Solutions

From the Literature

- Energy management with **no market prices** (Liu et al. [2021])
- Limited battery simulation with **no degradation costs** (Shojaeighadikolaei et al. [2021], Kolodziejczyk et al. [2021], Domínguez-Barbero et al. [2020], Chen and Bu [2019], Guo et al. [2022])
- Accurate battery dynamics, but **simplistic aging model** (Sui and Song [2020], Lin et al. [2021])

The Proposed Solutions

From the Literature

- Energy management with **no market prices** (Liu et al. [2021])
- Limited battery simulation with **no degradation costs** (Shojaeighadikolaei et al. [2021], Kolodziejczyk et al. [2021], Domínguez-Barbero et al. [2020], Chen and Bu [2019], Guo et al. [2022])
- Accurate battery dynamics, but **simplistic aging model** (Sui and Song [2020], Lin et al. [2021])
- Strong assumptions on energy prices, **limited overall realism** (Mussi et al. [2024])

The Proposed Solutions

From this Work

Main Contributions

1. Design of a **Reinforcement Learning**-based control strategy for microgrids
 - integrated with a **Digital Twin** for accurate simulation of battery dynamics
 - equipped with a **battery aging model** to account for degradation costs
 - responsive to **dynamic market prices** and **ambient temperature profiles**

The Proposed Solutions

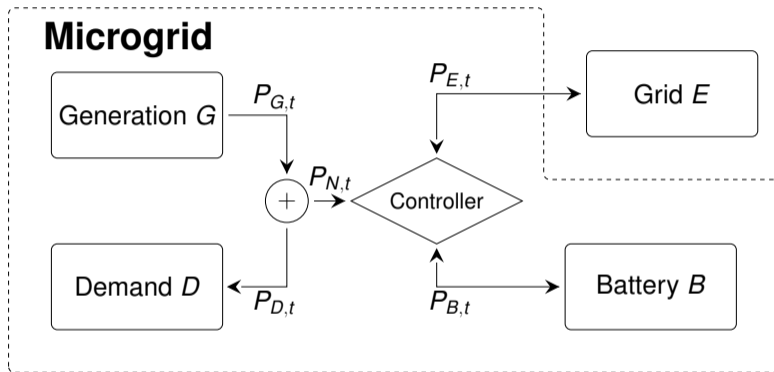
From this Work

Main Contributions

1. Design of a **Reinforcement Learning**-based control strategy for microgrids
 - integrated with a **Digital Twin** for accurate simulation of battery dynamics
 - equipped with a **battery aging model** to account for degradation costs
 - responsive to **dynamic market prices** and **ambient temperature profiles**
2. Validation on **real-world** energy datasets

Problem Formulation

The Microgrid Environment



$$P_{N,t} = P_{G,t} - P_{D,t}$$

$$P_{B,t} = a_t P_{N,t}$$

$$P_{E,t} = (1 - a_t) P_{N,t}$$

Problem Formulation

Objective Function

The objective R_T is the **maximization of the cumulative profit** over the time horizon T

Objective

$$R_T(\pi) = \sum_{t=1}^T [r_{\text{trad}}(a_t) + r_{\text{deg}}(a_t)]$$

- $r_{\text{trad},t}$ represents the **trading revenue** at time t
- $r_{\text{deg},t}$ is the **degradation cost** at time t ($r_{\text{deg},t} < 0$)
- π is the strategy adopted and a_t the corresponding action at time t

RL Problem Formulation

Markov Decision Process

MDP

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma, \mu_0)$$

RL Problem Formulation

Markov Decision Process

MDP

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \gamma, \mu_0)$$

- $s_t \in \mathcal{S}$:
 - internal battery state (state of charge, temperature)
 - energy demand / generation estimates
 - market prices
 - seasonal / time-of-day encodings
- $a_t \in \mathcal{A}$: percentage of $P_{N,t}$ addresses with the battery
- $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ reward function

$$r_t = r_{\text{trad}}(s_t, a_t) + r_{\text{deg}}(s_t, a_t) + \lambda r_{\text{clip}}(s_t, a_t)$$

Learning Procedure

Algorithm

Requirements of the algorithm:

- Profiles of demand, generation, market prices and ambient temperature
- A battery DT

Learning Procedure

Algorithm

Requirements of the algorithm:

- Profiles of demand, generation, market prices and ambient temperature
- A battery DT

For each step of the episode:

1. Estimate $P_{G,t}$ and $P_{D,t}$
2. Observe state s_t
3. Choose action $a_t \sim \pi(s_t)$
4. Compute $P_{B,t}$ and $P_{E,t}$
5. Update the virtual battery (simulated with the DT)
6. Collect reward r_t
7. Update policy π

Learning Procedure

Algorithm

Requirements of the algorithm:

- Profiles of demand, generation, market prices and ambient temperature
- A battery DT

For each step of the episode:

1. Estimate $P_{G,t}$ and $P_{D,t}$
2. Observe state s_t
3. Choose action $a_t \sim \pi(s_t)$
4. Compute $P_{B,t}$ and $P_{E,t}$
5. Update the virtual battery (simulated with the DT)
6. Collect reward r_t
7. Update policy π

Until convergence to π^*

Experimental Campaign

Experimental Setting

- Offline data (2015–2020):
 - 398 Italian household demand profiles
 - generation profiles from 3KW photovoltaic panel with Italian irradiation factor
 - Italian energy market prices
 - average Italian ambient temperature profiles
- Environment simulated with ErNESTO Digital Twin [Salaorni, 2023]

Experimental Campaign

Experimental Setting

- Offline data (2015–2020):
 - 398 Italian household demand profiles
 - generation profiles from 3KW photovoltaic panel with Italian irradiation factor
 - Italian energy market prices
 - average Italian ambient temperature profiles
- Environment simulated with ErNESTO Digital Twin [Salaorni, 2023]
- Offline RL agent based on PPO [Schulman et al., 2017]

Experimental Campaign

Baselines

Rule-based:

- OnlyGrid (OG)
- BatteryFirst (BF)
- 20-80, 80-20, 50-50

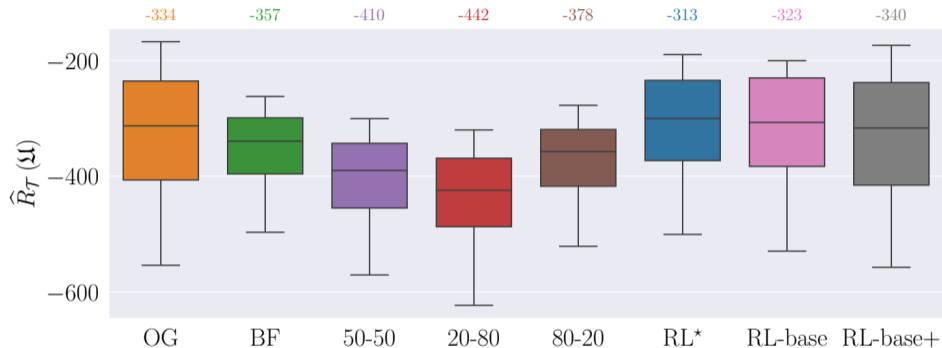
RL-based:

- RL-base: fixed energy prices, no ambient temperature (Mussi et al. [2024])
- RL-base+: dynamic prices, no ambient temperature

Our method: **RL***

Experimental Campaign

Return Comparison



- Paired t-test to check if the difference in rewards between RL^* and other methods is significantly greater than zero yielded an overall p-value of 0.0052, indicating strong statistical evidence that our method consistently outperforms the others

Conclusions

Outcomes of this work:

- Novel RL-based formalization of microgrid energy management
- Broad experimental campaign based on real-world data
- **RL*** outperforms state-of-the-art and rule-based controllers

Conclusions

Outcomes of this work:

- Novel RL-based formalization of microgrid energy management
- Broad experimental campaign based on real-world data
- **RL*** outperforms state-of-the-art and rule-based controllers

Future works:

- Application to a multi-agent setting (multiple microgrids, renewable energy communities, etc.)
- Design of specific MDP formulations for the presented setting

References

- Linpeng Liu, Jianquan Zhu, Jiajun Chen, and Hanfang Ye. Deep Reinforcement Learning for Stochastic Dynamic Microgrid Energy Management. In *2021 IEEE 4th International Electrical and Energy Conference (CIEEC)*, pages 1–6, 2021. doi: 10.1109/CIEEC50170.2021.9511049.
- Amin Shojaeighadikolaei, Arman Ghasemi, Alexandru G. Bardas, Reza Ahmadi, and Morteza Hashemi. Weather-Aware Data-Driven Microgrid Energy Management Using Deep Reinforcement Learning. In *2021 North American Power Symposium (NAPS)*, pages 1–6, 2021. doi: 10.1109/NAPS52732.2021.9654550.
- Waldemar Kolodziejczyk, Izabela Zoltowska, and Pawel Cichosz. Real-time energy purchase optimization for a storage-integrated photovoltaic system by deep reinforcement learning. *Control Engineering Practice*, 106:104598, 2021. ISSN 0967-0661. doi: <https://doi.org/10.1016/j.conengprac.2020.104598>.
- David Domínguez-Barbero, Javier García-González, Miguel A. Sanz-Bobi, and Eugenio F. Sánchez-Úbeda. Optimising a Microgrid System by Deep Reinforcement Learning Techniques. *Energies*, 13(11), 2020. ISSN 1996-1073. doi: 10.3390/en13112830.
- Tianyi Chen and Shengrong Bu. Realistic Peer-to-Peer Energy Trading Model for Microgrids using Deep Reinforcement Learning. In *2019 IEEE PES Innovative Smart Grid Technologies Europe*, pages 1–5, 2019. doi: 10.1109/ISGTEurope.2019.8905731.
- Chenyu Guo, Xin Wang, Yihui Zheng, and Feng Zhang. Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy*, 238:121873, 2022. ISSN 0360-5442. doi: <https://doi.org/10.1016/j.energy.2021.121873>.
- Yu Sui and Shiming Song. A multi-agent reinforcement learning framework for lithium-ion battery scheduling problems. *Energies*, 13(8), 2020. ISSN 1996-1073. doi: 10.3390/en13081982.
- S. H. Lin, H. H. Yu, and H. W. Chen. On-Line Optimization of Microgrid Operating Cost Based on Deep Reinforcement Learning. *IOP Conference Series: Earth and Environmental Science*, 701(1):012084, March 2021. doi: 10.1088/1755-1315/701/1/012084.
- Marco Mussi, Luigi Pellegrino, Oscar Francesco Pindaro, Marcello Restelli, and Francesco Trovò. A Reinforcement Learning controller optimizing costs and battery State of Health in smart grids. *Journal of Energy Storage*, 82, 2024. ISSN 2352-152X. doi: <https://doi.org/10.1016/j.est.2024.110572>.
- Davide Salaorni. Ernesto-dt, 2023. <https://github.com/Daveonwave/ErNEST0-DT>.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv*, 2017.

Thank you!

Questions?



**ErNESTO-gym: check
out this work**

Learning Procedure

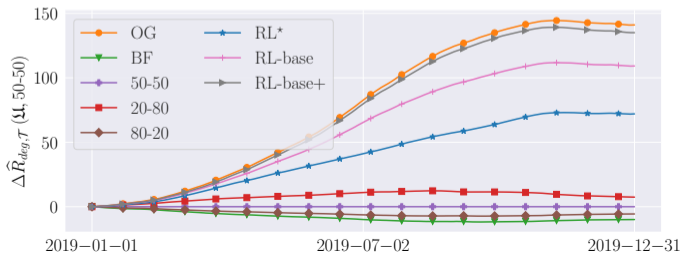
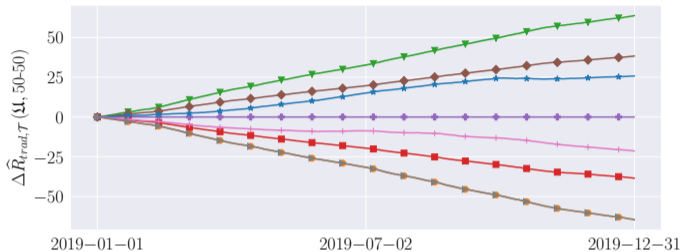
Pseudocode

Algorithm 1 Interaction between Agent and Environment

- 1: **Initialize:** $s_0, \{\mathcal{P}_D^{(i)}\}_{i=1}^M, \mathcal{P}_G, \mathcal{C}_{buy}, \mathcal{C}_{sell}, \mathcal{K}, B(\cdot), \pi(\cdot)$
 - 2: **for** $j \in \{1, \dots, n_{ep}\}$ **do**
 - 3: Sample demand profile $\mathcal{P}_D^{(i)} \sim Unif(\{\mathcal{P}_D^{(i)}\}_{i=1}^M)$
 - 4: Initialize σ_1, T_1, ρ_1
 - 5: **for** $t \in \{1, \dots, \mathcal{T}\}$ **do**
 - 6: Compute estimates $\hat{P}_{G,t}, \hat{P}_{D,t}^{(i)}$
 - 7: Observe current state s_t
 - 8: Agent takes action $a_t \sim \pi(s_t)$
 - 9: Compute $P_{B,t} \leftarrow a_t(P_{G,t} - P_{D,t}^{(i)})$
 - 10: Update $(\sigma_{t+1}, T_{t+1}, \rho_{t+1}) \leftarrow B(\sigma_t, T_t, K_t, P_{B,t})$
 - 11: Compute $P_{E,t} \leftarrow (1 - a_t)(P_{G,t} - P_{D,t}^{(i)})$
 - 12: Collect reward r_t
 - 13: Update policy $\pi(\cdot)$
-

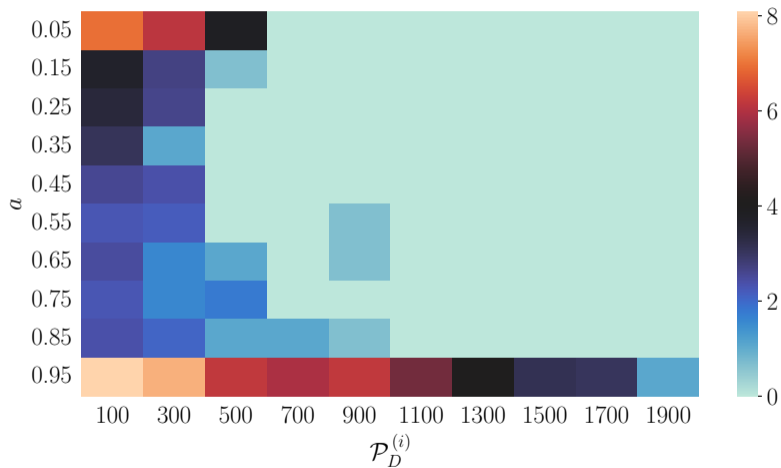
Experimental Campaign

Trading vs Degradation



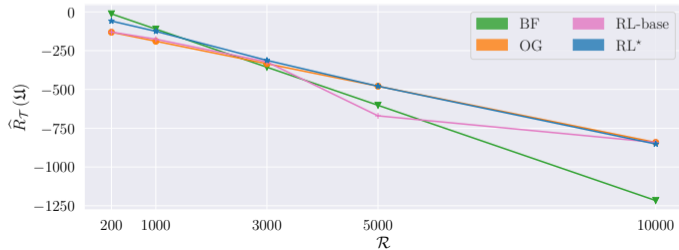
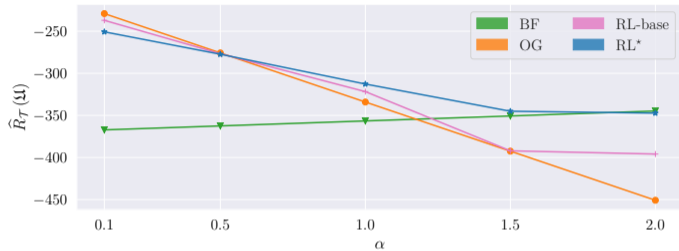
Experimental Campaign

Policy Behavior



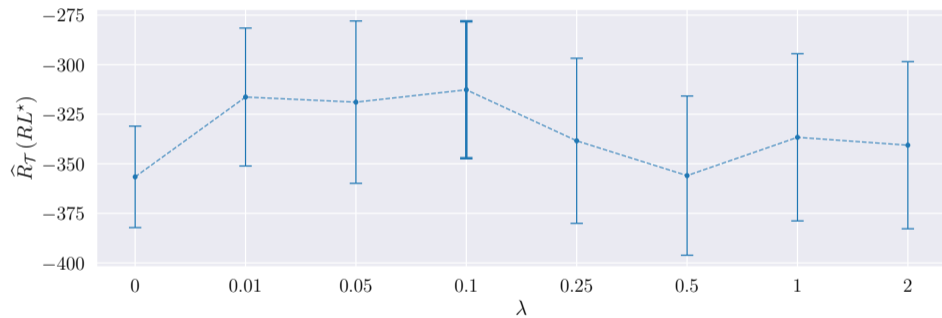
Experimental Campaign

Robustness Test



Experimental Campaign

Ablation Studies



Experimental Campaign

Training Hyperparameters

Parameter	Value
# Episodes	100
# Envs	8
Policy Network Size	[64, 32]
Gamma	0.99
Learning Rate	$5e-5$
Batch size	512
# Epochs	10
Rollouts	8912
Initial log. std.	-1
Normalize obs.	True
Seed	42