Trade-offs Between Profit and Sustainability in Multi-Agent PV Microgrid Energy Management: A Comparison of Profit-Profit Negotiation and Nash Q-Learning Approaches

Ricardo Vicente*
Instituto Superior Técnico, University of Lisbon

Abstract

This paper investigates the use of autonomous agents in photovoltaic (PV) microgrids to address key challenges such as variable weather conditions, energy storage optimization, market price fluctuations, system faults, and cybersecurity threats. A decentralized Multi-Agent System (MAS) is proposed to enhance energy management, strengthen system resilience, and enable effective fault response. The hybrid agent-based architecture integrates two core algorithms—Negotiation and Nash Q-learning—that govern energy flow and optimize cost efficiency in a competitive environment, while ensuring secure grid interaction, revenue generation, and reduced component wear.

CCS Concepts

- **Computing methodologies** → **Intelligent agents**; **Multi-agent systems**; *Cooperative artificial intelligence*; *Agent / discrete models*.

Keywords

PV Systems, Renewable Energy, Multi-agent Systems, Reactive Agents, Deliberative Agents, Hybrid Agents, Negotiation, Nash Q-learning, Minmax Algorithm

1 Introduction

This paper investigates the use of autonomous agents in photovoltaic (PV) microgrids to address key challenges such as weather variability, storage management, market volatility, faults, and cybersecurity. We propose a decentralized Multi-Agent System (MAS) that enhances system resilience, optimizes energy flow, and ensures secure grid interactions. To contextualize our approach, we begin by reviewing relevant literature on MAS applications in PV-based microgrids in Section 1. The following sections present the experimental setup, agent architecture, and simulation results based on two core algorithms—Negotiation and Nash Q-learning—as illustrated in Figure 1, demonstrating the effectiveness of the proposed system.

2 Related Work

Autonomous agents, typically structured as Multi-Agent Systems (MAS), represent a cutting-edge method for controlling complex microgrids, especially those with PV [Altin et al. 2023]. These agents bring **autonomy, reactivity, proactivity, and social interaction**, enabling distributed intelligence [Logenthiran et al. 2011]. PV-specific agents manage functions like Maximum Power Point Tracking (MPPT) and output coordination [Lopes et al. 2006; Pezeshki et al. 2011; Xiao et al. 2010], with Rahmani et al. [2017] and Azeroual et al. [2022] implementing component-specific agents, including solar generation and storage.

Key MAS applications include **energy optimization and management** [Azeroual et al. 2022; Logenthiran et al. 2011], resource scheduling [Logenthiran et al. 2011], and improving power quality and system stability [Altin et al. 2023], while also enabling fault handling [Coelho et al. 2017]. In islanded microgrids, MAS have proven effective in scheduling resources across microgrids and lumped loads using structured, multistage algorithms [Logenthiran et al. 2011].

Pipattanasomporn et al. [2012] developed a MAS for PV-based microgrids that can island during faults, prioritize critical loads, and resynchronize after recovery. Similarly, Rahmani et al. [2017] proposed an intelligent system to **reduce costs and operational disruption** by accounting for factors like backup generation and battery degradation.

Research has also explored Multi-Agent **Reinforcement Learning** (MARL) for managing renewable energy microgrids. Agents adapt their behaviour to optimize energy generation, storage, and operations [Nuvvula et al. 2024; Raju et al. 2015].

MAS are scalable and reconfigurable [Khan et al. 2018; Logenthiran et al. 2011], although **challenges** such as hardware integration and cybersecurity remain [Altin et al. 2023; Taleb et al. 2023], with dedicated security agents suggested as a solution [Taleb et al. 2023].

3 Environment

The environment of the present simulations includes the following key variables: the forecasted energy reaching the inverter (calculated based on the forecasted Global Tilted Irradiance [GTI] and the forecasted air temperature at 2 meters above ground level), forecasted market prices, the buy percentage of the available PV energy (including both panel-generated and battery-stored energy), PV faults (resulting from both environmental and technical causes), and cyberattacks entering the PV system through the grid. These variables together form a comprehensive dataset for simulating PV system behaviour under realistic conditions.

This project retrieves weather data every 15 minutes from May 1 to May 10, 2025, using the Open-Meteo API. Specifically, it utilizes the GTI (total irradiance) and the temperature at 2 meters height forecasts to calculate the **predicted**

*Student 113244, Group 14, AASMA 2025.

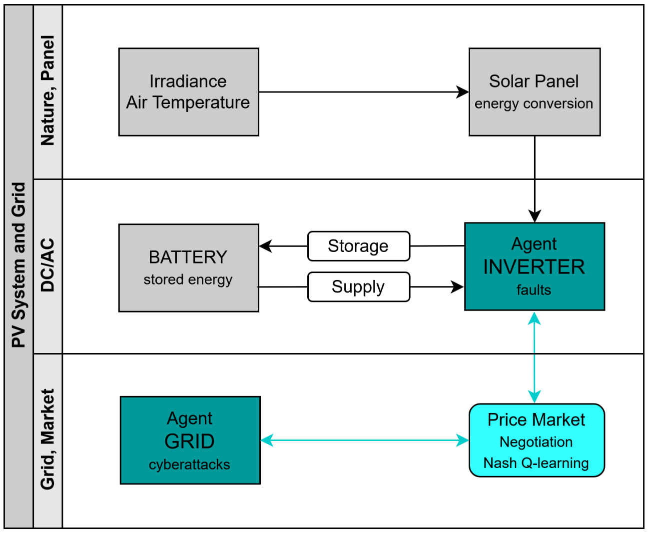

Figure 1: Diagram of the multi-agent system (Inverter/Battery and Grid) with environmental, financial, and technical variables

energy input to the inverter. This is done by first estimating the effective irradiance on the panel surface using the formula:

$$G_{\text{eff}} = \text{GTI} \cdot \cos\theta$$

where θ is the angle of incidence between the solar rays and the panel surface. A fixed incidence angle of 10° is used here, resulting in a cosine correction factor for the irradiance. Once the effective irradiance is known, the energy produced in direct current (EDC) is computed using the formula:

$$EDC = G_{\text{eff}} \cdot A_{\text{panel}} \cdot \eta_{\text{panel}} \cdot \Delta t$$

where A_{panel} is the total panel area (20000 m^2 in this setup), η_{panel} is the panel efficiency (assumed to be 20% under standard conditions), and Δt is the time interval of 15 minutes, expressed in hours ($\Delta t = \frac{15}{60} = 0.25$ h). This panel efficiency is also dynamically adjusted according to the temperature: it decreases by 0.5% for every degree Celsius above 25°C and increases by 0.4% per degree below 25°C.

The resulting EDC (in W) is then adjusted for transmission and conversion losses to compute the energy that actually reaches the inverter ($E_{\text{inverter\_input}}$). This is done by multiplying EDC by the cable efficiency (e. g., $\eta_{\text{cable}} = 98\%$) and the Maximum Power Point Tracking (MPPT) controller efficiency (e. g., $\eta_{\text{MPPT}} = 98\%$).

The energy **price forecast** is sourced from May 1 to May 10, 2025, from REN's official website. This data is originally provided in Excel format, which is manually edited and converted into a CSV file for processing. It includes timestamps and hourly prices in €/MWh.

A predefined **buy percentage** is used to model the portion of energy that needs to be bought from the PV system, based on EDP's three-period tariff scheme. These percentages are defined by time-of-day (Summer Time): 50% during off-peak hours (00:00–08:00 and 22:00–24:00), 75% during mid-peak hours (08:00–09:00, 10:30–18:00, and 20:30–22:00), and 100% during peak hours (09:00–10:30 and 18:00–20:30).

To simulate real-world conditions, the system also introduces **PV system faults**, as well as **cyberattacks** that can affect the PV system and infiltrate through the grid. These events are generated randomly on a daily basis, with 0 to 1 faults and 0 to 1 cyberattacks per day, each lasting between 15 and 60 minutes. During these periods, a binary flag is set to 1 to indicate a malfunction or attack, and 0 otherwise.

4 Energy Management

Battery management is explicitly modelled, highlighting the importance of energy storage in the system. In the setting of this research, a battery with a capacity of 200 kWh was used. The battery is represented with constraints on capacity limits, i. e. minimum and maximum thresholds for sustainable operation, and a **wear model** that calculates degradation cost over time based on energy usage. For simplification purposes, in this setting only battery discharging contributes to wear, which accumulates over time as an economic cost.

Battery wear is calculated during energy withdrawal as the product of discharged energy, current energy price, and a fixed wear coefficient (0.001). Notably, the standard battery storage limits are set between 20% and 80%. However, when the market price is significantly above average, the system allows the battery to discharge down to 10% or charge up to 90%, adapting dynamically to maximize economic benefit (Figure 2).

The inverter interacts closely with the battery, deciding how much energy to store or withdraw. This integration of battery dynamics into the learning framework adds realism and complexity to the energy trading strategy.

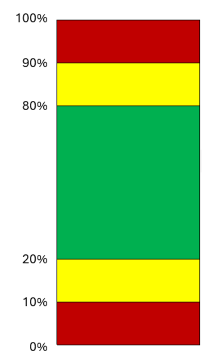

Figure 2: Upper and lower battery limits for storage and supply

Energy sold to the grid is reduced by DC-AC conversion losses (10%), and energy stored in the battery is subject to round-trip efficiency losses (3%). These technical losses are already deducted in the profit calculations, so the reported profits represent net gains from usable delivered energy.

Importantly, if **no trade** is made or a **cyberattack** is detected, the inverter attempts to store the unsold energy in the battery. However, only part of this energy can be stored, depending on available battery capacity. Any excess energy that cannot be stored is entirely lost, and its economic value is recorded as a loss. This models a realistic consequence in energy systems, where unused renewable energy is often wasted if not immediately stored or sold. Additionally, if a **fault** occurs in the PV system, the entire forecasted energy for that time step is lost.

5 Algorithms and Architecture

In this project, two algorithms were implemented: Negotiation Strategies and Nash Q-learning. The **Negotiation Strategies** algorithm includes multiple strategies, such as combinations of Fair, Yes-man, and Profit-oriented behaviours. For comparison with Nash Q-learning, the Profit-Profit strategy was chosen because it encourages competitive interactions between agents.

Nash Q-learning, on the other hand, is a reinforcement learning approach designed to handle multi-agent strategic interactions. To decide on actions within Nash Q-learning, a min-max algorithm was employed. This approach ensures that each agent selects actions that maximize its own expected payoff while anticipating the best possible responses from the opposing agent, effectively seeking a Nash equilibrium. This strategic reasoning allows agents to optimize their energy pricing decisions under uncertainty and competition.

Both the Profit-Profit negotiation strategy and Nash Q-learning aim to maximize profits through intelligent energy management and trading between the solar inverter owner and the grid operator. The agents use either pre-defined negotiation tactics or learned policies via Nash Q-learning with min-max action selection to optimize their gains through energy price setting.

Regarding the architecture, the codes of both simulations are primarily positioned in the **deliberative** category. Their agents do not simply react to immediate inputs but instead learn and plan using negotiation and Q-values. This deliberative behaviour is combined with **reactive** elements, such as adjusting decisions based on battery state, and detected faults or cyberattacks, resulting in an overall **hybrid architecture**.

The models were empirically fine-tuned to reduce lost energy and create a highly competitive interaction between the grid and inverter agents. Instead of using formal optimization, key parameters rate were adjusted through trial and error to better reflect realistic and competitive trading behaviour.

5.1 Negotiation Strategies

The simulation code for this algorithm is organized around **three core classes** that model the primary components of the energy negotiation system: the Inverter, the Battery, and the Grid. Battery and Inverter are both subclasses of the superclass **EnergyAgent** (Figure 3).

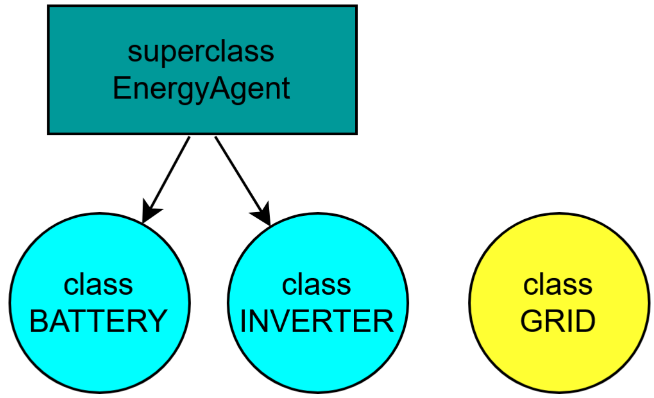

Figure 3: Class structure in the negotiation simulation

The Grid class models the external power grid and includes methods to negotiate prices with the inverter based on predefined negotiation modes (Fair, Yes-man, Profit-driven), receive energy, and track profits.

The Battery class manages energy storage, capacity limits, and wear due to discharging. It simulates realistic battery behaviour such as energy loss when withdrawing energy, and it tracks battery degradation over time.

The Inverter class dynamically negotiates prices with the grid, and controls energy flows between the panel, battery, and grid while accounting for efficiency losses. It also includes logic for adapting battery limits based on historical pricing, and deciding when to sell or store energy. This class also keeps track of profits and lost energy, adjusting its negotiation strategy based on modes similar to the grid.

The simulation function runs scenarios across all negotiation mode combinations, logs detailed time-step data, and summarizes performance metrics such as profits and energy losses, enabling comprehensive analysis of the negotiation dynamics.

Before a transaction is made, both the grid and inverter negotiate energy prices using the three possible behavioural modes: Fair, Yes-man, Profit-driven. The grid initiates offers below the market forecast price—by €10 (Yes-man), €5 (Fair), or €2 (Profit)—and gradually increases its bid until it reaches the forecast value, stepping by €1, €0.75, or €0.5 respectively. On the inverter side, it accepts any price in Yes-man mode, requires at least 40% of the forecast in Fair mode, and 60% in Profit-driven mode.

The inverter's **profit** is calculated as the amount of effective energy sold, after accounting for conversion and storage inefficiencies, multiplied by the agreed price. From the grid's perspective, profit comes from purchasing energy at a negotiated price lower than the market forecast. When the inverter accepts an offer, the grid calculates its profit as the difference between the forecast market price and the actual purchase price, multiplied by the energy effectively received. Hence, the cheaper the energy the grid can buy, the higher its profit margin.

Figure 4 shows the results for various negotiation strategy pairs (from Fair, Yes-man, and Profit).

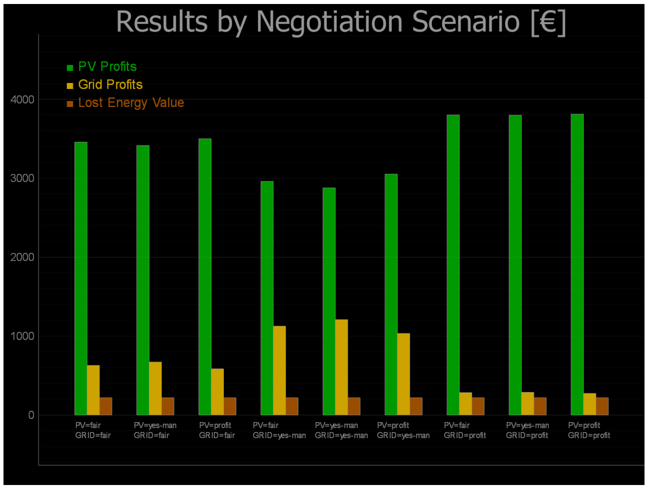

Figure 4: Results for negotiation pairs (Fair, Yes-man, and Profit)

Figure 5 illustrates the evolution of the environment and agent behaviours throughout the ten-day Profit-Profit negotiation simulation.

5.2 Nash Q-learning

At the top of this trading simulation, the abstract **EnergyAgent** superclass defines a general structure for all energy-related entities. Specific system components such as Battery and Inverter inherit from EnergyAgent, each extending and customizing the behaviour relevant to their roles: the Battery handles storage constraints and wear, the Inverter manages energy conversion and trading logic. The **NashQLearningAgent** superclass is particularly central, as it encapsulates the Nash Q-learning algorithm for both the inverter and grid agents, implementing the learning strategy for price management (Figure 6).

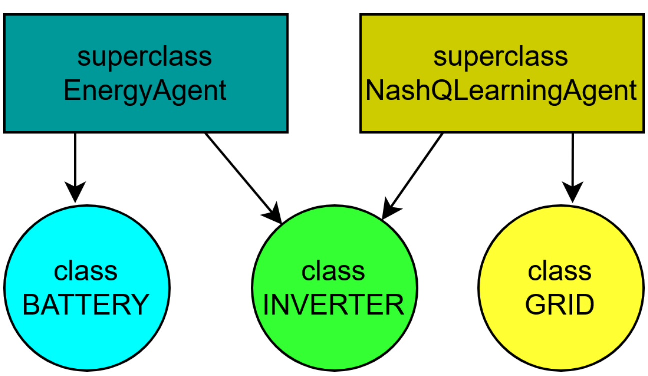

Figure 6: Class structure in the Nash Q-learning simulation

In this algorithm, **profit** is generated through an interaction between two agents: the Inverter and the Grid. Each agent's state includes forecasted energy input, battery level, and market price range, and choose prices (actions) from a unique price range. Each one strategically selects prices using Nash Q-learning, which considers the possible actions of the other agent and aims to maximize the agent's minimum guaranteed outcome (minmax algorithm). The total profit—calculated by multiplying the effective energy sold—is divided proportionally between the Inverter and the Grid based on their respective price offer (action).

Both the grid and inverter agents operate under partial observability, meaning that neither agent has access to the Q-table of the other. Each agent maintains its own Q-table, which maps environment states to expected returns for each action. During decision-making, an agent does not know how its opponent evaluates the same state or what the opponent's optimal strategy is. Instead, each agent assumes the opponent will choose actions uniformly at random across its available actions. This assumption reflects a competitive and decentralized learning setup, where agents are independently optimizing their strategies without coordination or direct knowledge of the opponent's learning process.

The main **parameters** of this simulation are presented below. These parameters, empirically fine-tuned, govern the learning behaviour and energy system dynamics within the simulation.

- **alpha = 0.1**: Learning rate. Determines how much new information overrides old knowledge in the Q-table.
- **gamma = 0.9**: Discount factor. Balances the importance of future values versus immediate ones.
- **epsilon = 0.2**: Exploration rate. Defines the probability of choosing a random action (exploration) rather than the best-known action (exploitation), being relatively high here.
- **q_initial = 0.1**: Initial Q-value. Sets a starting estimate for action.
- **reward = 3**: Absolute reward value assigned when a trade is successful (positive signal) or failed (negative signal).
- **states_energy = 5, states_battery = 5, states_price = 5**: Number of discrete levels used to represent forecasted energy, battery storage, and forecasted price. Thus, each state is represented as a tuple with three elements. In this scenario, there are $5 \times 5 \times 5 = 125$ total states.

If both agents' chosen prices are less than or equal to the forecasted price, the trade succeeds; otherwise, it fails. When the trade fails, no energy is sold to the grid—instead, all the forecasted energy is directed to battery storage, which may result in energy loss if the battery reaches its capacity limits.

The **reward policy** in this Nash Q-learning setup assigns a positive reward when both agents propose prices that are less than or equal to the forecasted price. In this case, they receive the same positive reward. If either agent proposes a price higher than the forecasted price, that agent receives

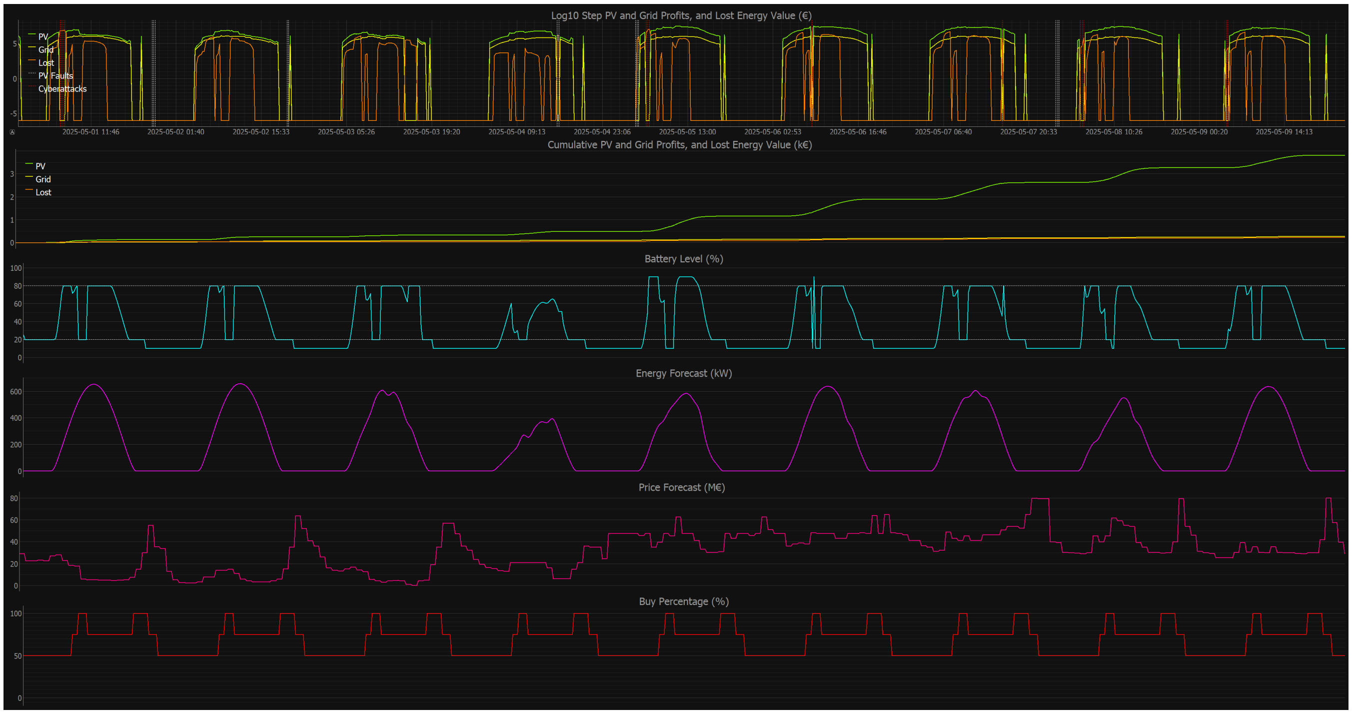

Figure 5: Complete simulation of a Profit-Profit negotiation

a negative reward, reflecting an uncompetitive offer, while the other agent still receives a positive reward symmetrical to that of the first. This design encourages both agents to select prices that are reasonable relative to the forecast and penalizes high prices that would prevent successful deals. The Q-values are updated accordingly to reflect these rewards and guide future price selections.

Figure 7 displays the evolution of the behaviour of the environment and the agents throughout the ten-day Nash Q-learning simulation.

6 Evaluating Metrics

Both simulations evaluate the behaviour of agents from a PV energy system—including an inverter and a battery—and an external grid, under different negotiation strategies. It does so by calculating several performance metrics that reflect how decisions about energy selling, storage, or loss affect both financial outcomes and system efficiency. These metrics are logged over time, both as cumulative totals and as values at each simulation step, enabling detailed analysis of short-term decisions and long-term trends.

Financial metrics are a key component. In the Negotiation Strategies algorithm, the inverter earns revenue by selling energy at a negotiated price, while the grid profits from the difference between the forecasted price and the negotiated price. In the Nash Q-learning framework, when trading occurs, the total revenue from selling energy is split between the inverter and the grid based on their price bids; the profit is divided proportionally, with each agent's share determined by its price offer relative to the combined bids of both agents.

Another important area is **energy efficiency**. The system tracks energy losses—specifically, beyond the losses from battery storage inefficiencies and DC-AC conversion, it also accounts for the energy produced by the PV system that cannot be sold or stored due to constraints such as a full battery or faults. Monitoring these values highlights inefficiencies in the system and helps identify potential areas for improvement.

Figure 8 presents the results for inverter and grid revenues, and the value of lost energy.

Figure 8: Inverter and Grid Revenues, and Lost Energy Value for Profit-Profit Negotiation and Nash Q-learning

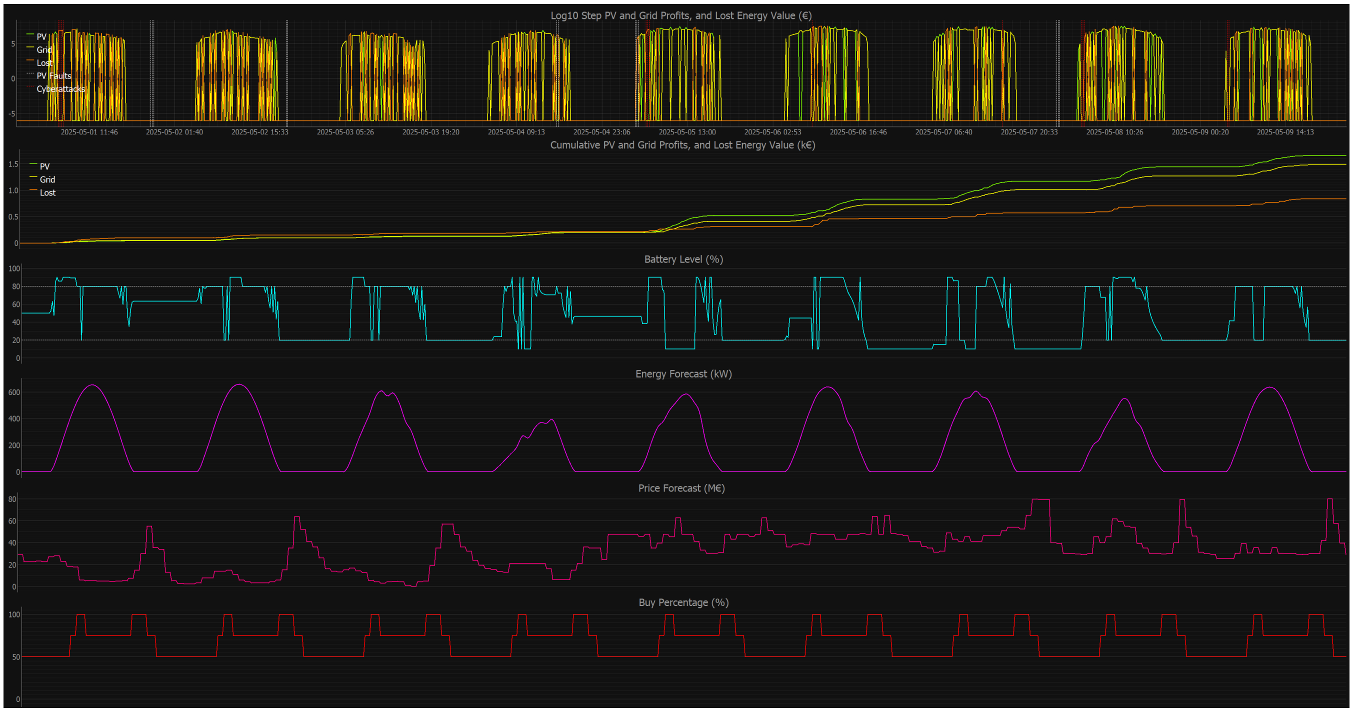

Figure 7: Complete simulation of a Profit-Profit negotiation

Battery wear is also calculated to assess long-term system resilience. The inverter tracks battery degradation based on how much energy is withdrawn and the price at the time, with the idea that higher-value charges and discharges contribute more to wear. Figure 9 presents the results for comparative battery wear.

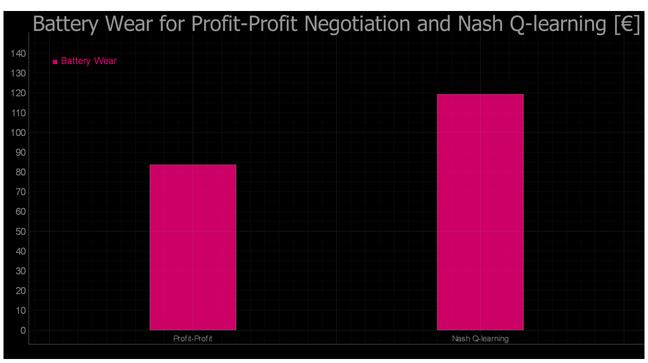

Figure 9: Battery wear for Profit-Profit Negotiation and Nash Q-learning

Combined, these metrics provide a comprehensive view of how negotiation strategies impact system performance, profitability, and sustainability.

7 Results Analysis

When comparing the Profit-Profit negotiation algorithm with Nash Q-learning, significant differences can be observed across four key performance indicators: inverter profit, grid profit, lost energy value, and battery wear. These indicators reveal not only how each agent performs economically, but also how efficiently the system uses energy and how intensively it relies on battery storage.

In the Profit-Profit **negotiation** scenario, the inverter achieves a much higher **profit** than the grid, indicating a negotiation dynamic that heavily favours the inverter. This **imbalance** suggests that the negotiation strategy used here lacks a fairness constraint or a mechanism that ensures mutual benefit. Instead, the inverter appears to dominate the negotiation, securing better trade terms. On the other hand, **Nash Q-learning** produces more **balanced** profits between the inverter and the grid, which aligns with the theoretical foundations of Nash equilibrium. In this approach, both agents adaptively learn policies that anticipate each other's decisions, resulting in outcomes where neither agent can unilaterally improve its performance—hence the observed profit symmetry.

Another interesting contrast is seen in the **lost energy value**, which is significantly **lower** in the **negotiation** setup compared to Nash Q-learning. This suggests that the inverter in the Profit-Profit scenario is more aggressive in capturing energy opportunities, likely by prioritizing immediate profits over longer-term considerations such as energy balancing or cooperation. The lower losses indicates that energy is being dispatched quickly to the grid, minimizing waste at the cost of system equity.

Battery wear presents another major difference. In the **negotiation** algorithm, battery wear is approximately two-thirds of that observed in Nash Q-learning. This suggests that the inverter in Profit-Profit uses the battery more conservatively, favouring grid trading over storing energy. Such a strategy **reduces battery degradation**, which can be economically advantageous in the long term. Conversely, **Nash Q-learning** likely involves more frequent or deeper charge-discharge cycles, as the agents optimize over longer horizons and anticipate each other's moves. While this approach promotes balanced profits and a more cooperative dynamic, it comes at a cost: **accelerated battery degradation**, which negatively impacts the PV system. Over time, this can significantly shorten the battery's operational lifespan, increasing maintenance expenses and potentially necessitating earlier replacement. This highlights a key limitation of fairness-focused strategies—they may be more equitable and energy-efficient in the short term, but their sustainability can be compromised by long-term hardware wear. Therefore, when evaluating these algorithms, it is crucial to consider not only immediate performance metrics but also long-term system viability and asset preservation.

8 Conclusions

Profit-Profit negotiation encourages short-term, high-yield strategies for the inverter, resulting in lower system losses and reduced battery stress, but this comes at the expense of equity between the agents. In contrast, Nash Q-learning fosters a highly competitive yet more balanced setting, where both agents achieve comparable profits. However, this balance is accompanied by slightly higher energy losses and increased battery wear, reflecting the more dynamic and adaptive interplay between agents.

These contrasting outcomes underscore the inherent trade-offs in multi-agent energy management: prioritizing immediate profit and system efficiency may reduce fairness and asset longevity, while striving for equitable gains can increase operational costs and hardware degradation. Ultimately, the choice between these approaches depends on the specific priorities of the energy system—whether the focus is on maximizing short-term returns or ensuring long-term sustainability and fairness.

The Profit-Profit negotiation model is well-suited for scenarios where short-term efficiency and minimizing system wear are prioritized, such as in microgrids with limited battery capacity or where operational costs must be tightly controlled. It benefits systems that can accept some imbalance between agents in exchange for reduced energy losses and longer battery life. On the other hand, the Nash Q-learning approach is ideal for contexts demanding fairness and adaptability, such as larger, more complex energy markets or multi-stakeholder environments where maintaining equitable profit distribution and dynamic negotiation is crucial. Although it may incur higher energy losses and battery wear, this model supports long-term collaboration and resilience by balancing agent interests more effectively.

References

Necmi Altin, Süleyman Emre Eyimaya, and Adel Nasiri. Multi-agent-based controller for microgrids: An overview and case study. *Energies*, 16(5):1–19, March 2023. doi: 10.3390/en16052445. URL https://doi.org/10.3390/en16052445.

Mohamed Azeroual, Younes Boujoudar, Lahcen EL Iysaouy, Ayman Aljarbouh, Muhammad Fayaz, Muhammad Shuaib Qureshi, Fazle Rabbi, and Hassane EL Markhi. Energy management and control system for microgrid based wind-pv-battery using multi-agent systems. *Wind Engineering*, 46(4):1247–1263, February 2022. doi: 10.1177/0309524X221075583. URL https://journals.sagepub.com/doi/10.1177/0309524X221075583.

Vitor N. Coelho, Miri Weiss Cohen, Igor M. Coelho, Nian Liu, and Frederico Gadelha Guimarães. Multi-agent systems applied for energy systems integration: State-of-the-art applications and trends in microgrids. *Applied Energy*, 187(6):820–832, February 2017. doi: 10.1016/j.apenergy.2016.10.056. URL http://dx.doi.org/10.1016/j.apenergy.2016.10.056.

Muhammad Waseem Khan, Jie Wang, Linyun Xiong, and Meiling Ma. Modelling and optimal management of distributed microgrid using multi-agent systems. *Sustainable Cities and Society*, 41:154–169, 2018. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2018.05.018. URL https://www.sciencedirect.com/science/article/pii/S2210670718302312.

T. Logenthiran, Dipti Srinivasan, and Ashwin M. Khambadkone. Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system. *Electric Power Systems Research*, 81(1): 138–148, January 2011. doi: 10.1016/j.epsr.2010.07.019. URL https://doi.org/10.1016/j.epsr.2010.07.019.

J.A.P. Lopes, C.L. Moreira, and A.G. Madureira. Defining control strategies for microgrids islanded operation. *IEEE Transactions on Power Systems*, 21(2):916–924, 2006. doi: 10.1109/TPWRS.2006.873018. URL https://ieeexplore.ieee.org/document/1626398.

Ramakrishna S S Nuvvula, Polamarasetty P Kumar, Syed Riyaz Ahammed, Vanam. Satyanarayana, Bachina Harish Babu, R. Siva Subramanyam Reddy, and Ahmed Ali. Distributed multi-agent reinforcement learning for autonomus management of renewable energy microgrids. In *2024 12th International Conference on Smart Grid (icSmartGrid)*, pages 298–304, 2024. doi: 10.1109/icSmartGrid61824.2024.10578150.

H. Pezeshki, P.J. Wolfs, and M. Johnson. Multi-agent systems for modeling high penetration photovoltaic system impacts in distribution networks. In *2011 IEEE PES Innovative Smart Grid Technologies*, pages 1–8, 2011. doi: 10.1109/ISGT-Asia.2011.6167149. URL https://ieeexplore.ieee.org/document/6167149.

Manisa Pipattanasomporn, Hassan Feroze, and Saifur Rahman. Securing critical loads in a pv-based microgrid with a multi-agent system. *Renewable Energy*, 39(1):166–174, March 2012. doi: 10.1016/j.renene.2011.07.049. URL https://doi.org/10.1016/j.renene.2011.07.049.

R. Rahmani, I. Moser, and M. Seyedmahmoudian. Multi-agent based operational cost and inconvenience optimization of pv-based microgrid. *Solar Energy*, 150:177–191, 2017. ISSN 0038-092X. doi: https://doi.org/10.1016/j.solener.2017.04.019. URL https://www.sciencedirect.com/science/article/pii/S0038092X17303043.

Leo Raju, Sibi Sankar, and R.S. Milton. Distributed optimization of solar micro-grid using multi agent reinforcement learning. *Procedia Computer Science*, 46:231–239, 2015. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2015.02.016. URL https://www.sciencedirect.com/science/article/pii/S1877050915000800. Proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014 at Bolgatty Palace & Island Resort, Kochi, India.

Ihab Taleb, Guillaume Guerard, Frédéric Fauberteau, and Nga Nguyen. A holonic multi-agent architecture for smart grids. In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART*, pages 126–134. INSTICC, SciTePress, 2023. ISBN 978-989-758-623-1. doi: 10.5220/0011803300003393. URL https://www.scitepress.org/Link.aspx?doi=10.5220/0011803300003393.

Zhe Xiao, Tinghua Li, Ming Huang, Jihong Shi, Jingjing Yang, Jiang Yu, and Wei Wu. Hierarchical mas based control strategy for microgrid. *Energies*, 9(3):1622–1638, September 2010. doi: 10.3390/en3091622. URL https://doi.org/10.3390/en3091622.