# Towards DQN Reinforcement Learning for energy management for bidirectional charging of EV's

Rainer Gasper[1], Michael Quarti[1], Nick Abermeth[1], Yannik Heizmann[1], Joshua Ruf[1], Markus Portugal[1], Bennet Märtin[1]

[1] Hochschule Offenburg
**rainer.gasper@hs-offenburg.de**

**Abstract.** This study applies Deep Q-Network (DQN) reinforcement learning to optimize bidirectional EV charging in a microgrid with dynamic pricing and renewable energy. The environment includes an EV, wind turbine, stationary battery, flexible household loads, and grid connection. DQN agents learn to minimize energy costs by charging during low-price periods and discharging during high-price windows. Simulations across four scenarios show improved cumulative rewards and grid efficiency. Future work will address stochastic elements, realistic EV availability, and continuous action spaces to enhance adaptability and performance in real-world applications.

**Keywords:** Reinforcement Learning; DQN; MicroGrid, Bidirectional Charging.

## 1 Introduction

The integration of electric vehicles (EVs) into microgrids presents a transformative opportunity for local energy systems. The bidirectional charging of EVs enables vehicle-to-grid (V2G) services, which leads to enhancing grid stability, reducing peak loads, and supporting renewable energy integration. However, managing bidirectional charging in dynamic environments like microgrids is complex due to fluctuating energy demand, supply, and user behavior.

Deep Q-Network (DQN) reinforcement learning offers a promising strategy to address this challenge. By learning optimal policies through interaction with the environment, DQN can manage charging and discharging decisions, adapting to real-time conditions and maximizing long-term rewards such as cost savings.

## 2 DQN Reinforcement Learning

DQN combines Q-learning with deep neural networks to approximate the optimal policy function. The DQN is implemented with experience replay, where the transitions (state, action, reward, next state) are stored in a memory buffer and sampled randomly during training [2]. Also, a target network is used to compute target values [3]. Both technics improve the learning process by reducing oscillations and breaking correlation between sequential data [2,3]. For the exploitation and exploring trade-off, we implemented Ɛ-greedy-strategy with decreasing Ɛ. The Q-learning allows only discrete actions (e.g. charge, discharge, buy or sell) based on the current state of the environment. The learning takes place in the agent, he receives rewards based on its actions and updates its policy to maximize cumulative future rewards, the cost savings.

The environment consists of the microgrid and the reward function. The microgrid itself consists of an EV with bidirectional charging capability, a wind turbine as a renewable energy source, a stationary battery, and households as consumers with flexibilities to shift their power consumption [3]. Also, the grid connection point with the dynamic electricity pricing is included, which acts as a source (buying) or sink (selling). The data for the microgrid are taken from [3]. We neglected the uncertainties in the microgrid, therefore the environment is fully deterministic. The reward function calculates the energy costs (multiplied by minus one).


**Fig. 1.** Test results for net profit over 100 days for baseline batterie, baseline grid, price action and shift action

## 3 Simulation Results

In the simulation, we compare four different scenarios to evaluate the results. The first scenario is the baseline, where no batteries are used and every excess power is sold, and vice versa. The second scenario includes the stationary battery, where the excess power is stored in the batterie, and a power gap leads to discharging of the battery. Only if the limits of the battery are reached, the grid power is used. In the third scenario the agent minimized the price, and in the fourth scenario, the agent can also shift the power demand of some consumers. The initial results, see Fig. 1, for 150 households show that the agents learn to discharge during high-price periods and charge during low-price windows. Compared to baseline scenarios, the DQN-agents achieve higher cumulative rewards.

## 4 Conclusion and Outlook

The first results show already consistent behavior of the DQN – agents. Future work will also include stochastics in the renewable energy, the price forecast, and in the consumer behavior. This includes the availability of the EV battery. Also, the influence of different reward functions has to be investigated further. Another point for future work is to implement agents with continuous actions for power distribution.

## References

1. T. A. Nakabi, P. Toivanen, Deep reinforcement learning for energy management in a microgrid with flexible demand, Sust. Energy, Grids and Networks, Volume 25, 2021, https://doi.org/10.1016/j.segan.2020.100413

2. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., et al. (2013). Playing Atari with Deep Reinforcement Learning. CoRR, abs/1312.5602.
3. Mnih, V., Kavukcuoglu, K., Silver, D. et al., Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). https://doi.org/10.1038/nature14236