

# DISTRIBUTED REINFORCEMENT LEARNING FOR DC OPEN ENERGY SYSTEMS

**Qiong Huang & Kenji Doya**

Neural Computation Unit, Okinawa Institute of Science and Technology Graduate University  
Okinawa 9040412, Japan  
{qiong.huang, doya}@oist.jp

## ABSTRACT

The direct current open energy system (DCOES) enables the production, storage, and exchange of renewable energy within local communities, which is helpful, especially in isolated villages and islands where centralized power supply is unavailable or unstable. As solar and wind energy production varies in time and space depending on the weather and the energy usage patterns differ for different households, how to store and exchange energy is an important research issue. In this work, we explore the use of deep reinforcement learning (DRL) for adaptive control of energy storage in local batteries and energy sharing through DC grids. We extend the Autonomous Power Interchange System (APIS) emulator from SonyCSL to combine it with reinforcement learning algorithms in each house. We implemented deep Q-network (DQN) and prioritized DQN to dynamically set the parameters of the real-time energy exchange protocol of APIS and tested it using the actual data collected from the DCOES in the faculty houses of Okinawa Institute of Science and Technology (OIST). The simulation results showed that RL agents outperformed the hand-tuned control strategy. Sharing average energy production, storage, and usage within the local community further improved efficiency. The implementation of DRL methods for adaptive energy storage and exchange can help reducing carbon emission and positively impact the climate.

## 1 INTRODUCTION

A local DC-based microgrid system was set up at the OIST faculty housing area in Okinawa, Japan (Tokoro, 2014; Werth et al., 2015). It focuses on redefining electricity grid systems in the form of interconnected nano-grid subsystems. It is designed to harness a mixture of renewable energy sources, including photovoltaic (PV) panels, rechargeable batteries, and local DC grid. It is also designed to be deployed in various communities and is especially helpful in rural, isolated communities. In contrast to conventional centralized energy systems, the distributed design of DCOES allows the energy source close to consumers with short electric mileage and energy storage in batteries to fill the temporal gap between supply and demand (Sakagami et al., 2015). Each house is equipped with PV panels and an energy storage system (ESS), including lithium-ion batteries, through which it is connected to the others via DC networks (Tokoro, 2015). Figure 1 is a sketch of energy exchange among several houses. Houses are connected with the energy exchange DC Grid Bus Line (blue dashed line) and a Communication Line (orange dashed line) to determine the deals of request/accept energy. With its ability to request and receive energy from the local neighbor nodes, this system has been proven in a real environment to be efficient in the use of renewable energy, flexible in size, as well as to increase dependability when used with utility electricity (Sakagami et al., 2016).

So far, various approaches for optimizing the energy exchange have been applied to the DCOES. (Werth et al., 2016) devised a peer-to-peer control for DC microgrids. They found that PV and battery with DC exchanges outperformed the PV-only (classic home system) cases and the standalone (without DC exchanges) cases and performed close to the theoretical limit of centralized control. In addition, (Sakagami et al., 2016) also compared the standalone cases and DCOES case. They calculated the DCOES self-sufficiency rate (DSSR) to evaluate the performances in each case and found the optimal configurations of PV and installation cost. They found that the cost recovery

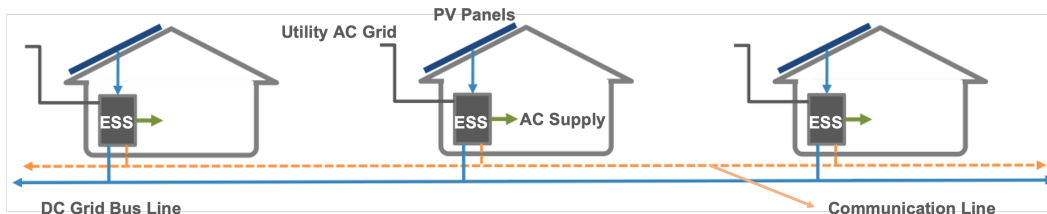


Figure 1: ESS system and DCOES connections among different houses.

period of the DCOES can further be cut down if more PVs and batteries are installed, while the current recovery period is 15 years Sakagami et al. (2016). More recently, (Kawamoto & Rajendiran, 2021) showed that using machine learning techniques could further optimize the energy exchange by minimizing the wasting of the surplus of solar energy. They found simple RNN (Elman-net) outperforms linear regression and LSTM methods in predicting the future hourly consumption from the past days to maintain the charging space for the battery. The battery could reserve more charging space and reduce the time of wasting surplus energy after the battery gets fully charged. The current control policy of DCOES is limited in using a fixed rule-based policy for energy sharing, which is manually adjusted. It is challenging to fine-tune for different households due to varying usage and storage levels. This results in a waste of energy. To further improve the effectiveness of the energy grid and achieve better energy management, it is desired to utilize an AI-based adaptive control system for energy sharing.

Energy management is a complex task that requires balancing the demand and supply of energy to ensure optimal utilization of resources and minimizing costs. Reinforcement learning is a promising approach for energy management to make decisions in the face of uncertainty through interaction with the environment. Many previous works also tried applying RL to these problems.

Hau et al. (2020) used a value-based reinforcement learning and DQN-based energy management algorithm for energy trading. They applied a piece-wise utility function in response to the dynamic environment under dynamic pricing. François-Lavet et al. (2016) applied DRL to MG with PV panels. They introduced a particular deep learning architecture to draw knowledge from historical time series of consumption and production as well as accessible forecasts. For peer-to-peer trade pricing, Zhou et al. (2019) used RL to represent the suggested energy trading process as an MDP and identify the best decision inside the MDP. Many reinforcement learning methods are typically designed for single agents, which have a suboptimal performance for multi-agent systems, neither for cooperative nor competitive environments. How to tackle the problem in the multi-agent scenario is also challenging. Many algorithms centrally train the strategies of all agents, and after the training, agents can have the ability to make distributed decisions, such as COMA in (Foerster et al., 2018) and MADDPG method in (Lowe et al., 2017). However, due to the centralized training process, it also has a certain scalability problem.

## 2 REINFORCEMENT LEARNING METHODS IN APIS

We modeled energy sharing in the DC grid as a multi-agent reinforcement learning problem. In this paper, we test multiple reinforcement learning algorithms and investigate what type of state, action, and reward representations are helpful for practical performance.

The APIS is an open-source power interchange management software developed by SonyCSL (Sony Computer Science Laboratories, Inc., 2021), which comprises the node software for P2P power interchange, the main controller for monitoring and visualization, and emulators of the DCOES hardware, including ESS. We use the APIS for energy-sharing experiments. The energy exchange rule for each house is defined by a scenario file, which is defined independently for each node. The deal negotiation is carried out as each node compares its battery's relative state of charge (RSOC) with the scenario information. Once the deal is made, energy exchange starts between the subsystems for deal execution.

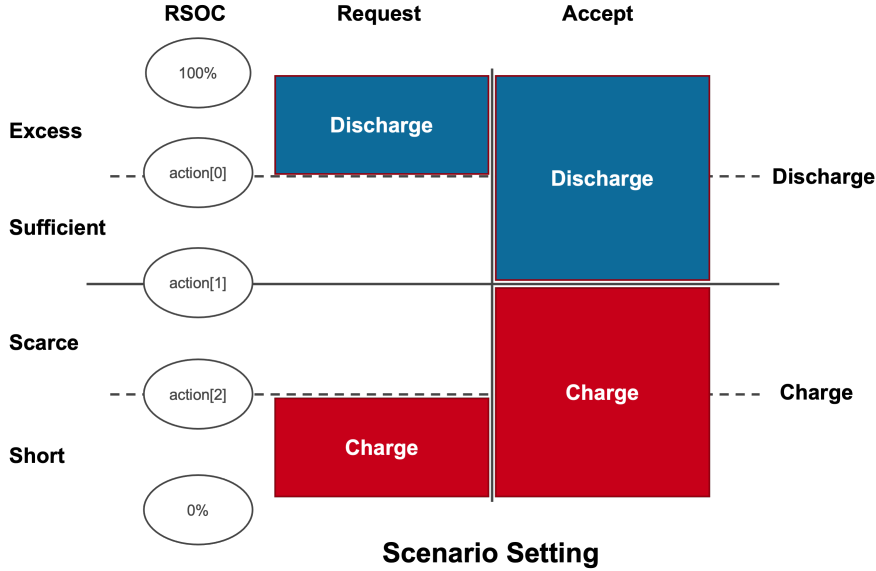


Figure 2: Action selection according to the RSOC in example scenario.

### 2.1 STATES, ACTION, AND REWARD SETTINGS

**States** In the multi-house setting, we tested three settings of state variables.

- 1) **Stand alone**  $\mathcal{S} = \{s_{pv}, s_{load}, s_{rsoc}, s_{p2}\}$
- 2) **With Community average**  $\mathcal{S} = \{s_{pv}, s_{load}, s_{rsoc}, s_{p2}, s_{rsoc_{ave}}, s_{ig}\}$
- 3) **With time of the day information**  $\mathcal{S} = \{s_{pv}, s_{load}, s_{rsoc}, s_{p2}, s_{rsoc_{ave}}, s_{ig}, s_{time}\}$ ,

where  $s_{pv}$  denotes photovoltaic power production,  $s_{load}$  denotes power consumption in the house,  $s_{rsoc}$  denotes the RSOC of the battery,  $s_{p2}$  is input power purchased from the utility grid,  $s_{rsoc_{ave}}$  denotes the community average RSOC,  $s_{ig}$  is the exchange grid current through the DC grid, and  $s_{time}$  is a two-dimensional encoding of daily cyclic with sine and cosine components.

**Action** We use an indirect control of energy exchange by setting the RSOC thresholds of the scenario file of APIS. In each time period, the battery status is classified into one of four levels: *Excess* above action[0], *Sufficient* above action[1], *Scarce* above action[2], and *Short* below (see Figure 2). A node in *Excess* status sends a Discharge-REQUEST, while a node in *Scarce* status sends a Charge-REQUEST to all nodes in the cluster. A node in *Excess* or *Sufficient* status sends Discharge-ACCEPT, while a node in *Sufficient* or *Scarce* status sends a Charge-ACCEPT. We used RSOC thresholds in discrete values in every 10% step from 20% to 90%. with a constraint of  $action[0] > action[1] > action[2]$ .

**Reward** The reward is formulated from the consumption power of the house. We tried two different reward settings, one using the sum of purchased power of the community and the other one using individually purchased power ( $pp_i$  where  $i$  denotes agent  $i$ ). The reward for each agent  $i$  is calculated from

$$reward_i = -pp_i,$$

and

$$reward_i = -\sum_i pp_i.$$

## 3 EXPERIMENTS

We implemented both DQN (Mnih et al., 2013) and Prioritized DQN (Schaul et al., 2015) methods to the APIS for 4 agents using data from May 8 to Jun 6 (30 days) in 2019. We update the action every

3 hours to change the conditions for deals between houses. Then we calculated the daily value of the purchased power and exchanged power, as well as the self-sufficiency rate (SSR), which is a value calculated as (Sakagami et al., 2015), among 4 agents.

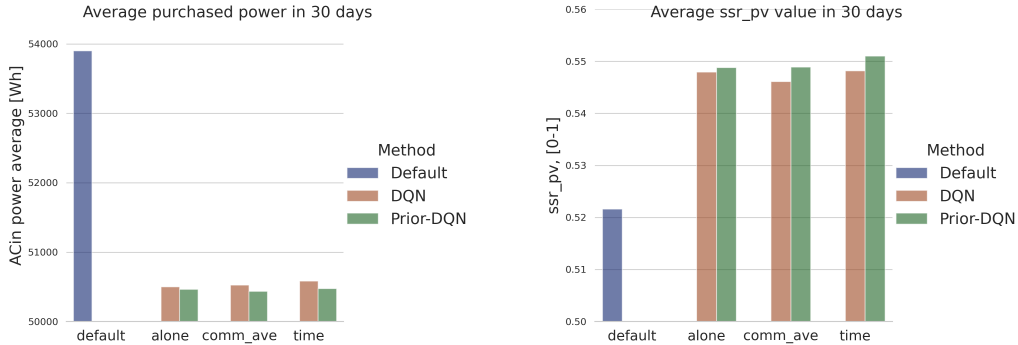
$$SSR = \frac{E_{self}}{E_{consumption}},$$

where  $E_{self}$  and  $E_{consumption}$  are, respectively, the electric power supplied locally by the PV panel, battery and DC exchange, and the total power consumed. The SSR is 100% when all the electricity consumed in the community is supplied by solar power. And SSR is 0% when electric power is only supplied by the external power supply. SSR is measured for each day.

#### 4 RESULTS AND DISCUSSION

Figure 3 shows the training results in DQN and prior-DQN methods compared with the fixed-rule control (default) case. Both DQN (the orange bar) and prior-DQN (the green bar) methods outperform the default case in buying electricity from the external power supply company (Figure 3a). For different state representations, using the average battery RSOC of the community has the least purchased energy, while having time-of-day information is the second, and stand-alone case is the last. Figure 3b shows the average SSR value across 30 days of each method. SSR values in both RL methods increased compared to the default rule-based control. In addition, prior-DQN outperforms DQN method in less purchase of energy and higher SSR value.

The results of prior-DQN learning with different reward settings are presented in Figure 4. As shown in Figure 4a, the sum reward case (the orange bar) outperforms the individual reward case (the green bar). Figure 4b demonstrates that using the sum reward function can lead to a higher average SSR value than using individual reward settings.



(a) Average daily purchased power, different methods.

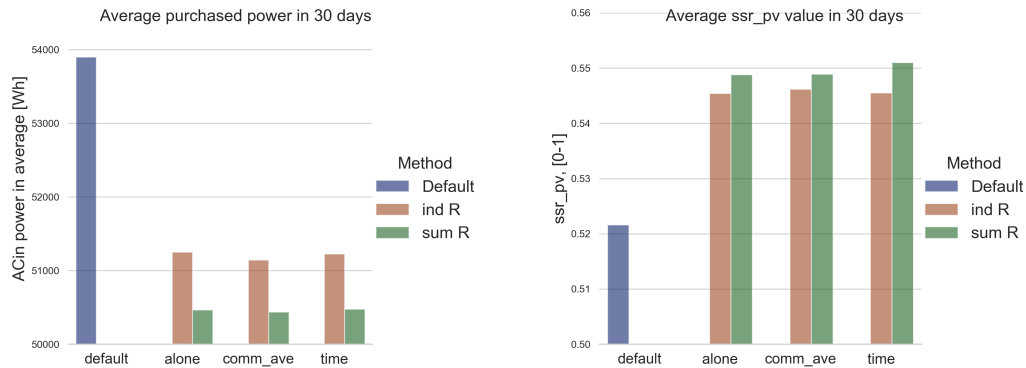
(b) Average SSR values, different methods.

Figure 3: Performance of different state options in different control methods.

In addition, we also compared the exchanged power in different cases. Figure 5a shows the average exchanged power with DQN and Prior-DQN methods. It is clear that compared with the default case, both DRL methods have more exchanged power. Prior-DQN agents further have more exchanged power comparing with DQN agents. Figure 5b reveals that using the sum reward function also has more exchanged power than individual reward.

#### 5 CONCLUSION AND FURTHER WORK

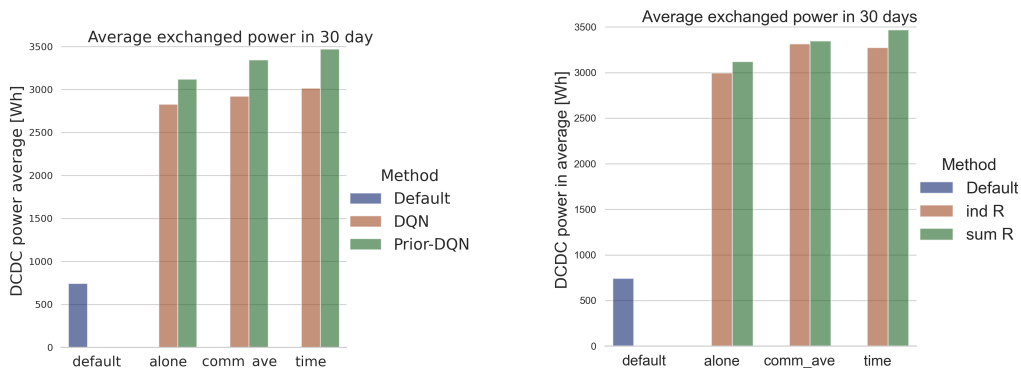
In this paper, we implemented DRL methods in APIS of DCOES for multiple-house energy management. The simulation results indicate that DRL methods outperform the rule-based control of energy sharing. The prior-DQN method is better than DQN in reducing the external purchase of electricity, improving the SSR value, and increasing the exchanged power. Using the community average RSOC can further improve performance. Moreover, using the sum reward of the community outperform the



(a) Average daily purchased power, different reward settings.

(b) Average daily SSR value with different reward settings.

Figure 4: Performance of different states options in different reward settings with prior-DQN method.



(a) Average daily exchanged power with different methods.

(b) Average exchanged power, different reward settings.

Figure 5: Exchanged power with different states options of different methods and different reward settings.

use of the individual reward of the agent in improving grid efficiency. The results indicate that DRL approaches can decrease the demand for fossil fuel-generated power by optimizing energy usage and reducing the external purchase of electricity. Additionally, by improving the efficiency of the energy grid, DRL methods can reduce the need for additional power infrastructure, which can further reduce the carbon footprint of energy production and distribution. The result in this paper is from one iteration in each experiment, and our further work will focus on multiple runs and iterations to verify the robustness of the learning, the generalization of the trained model to different season data and other multi-agent reinforcement learning methods.

## ACKNOWLEDGMENT

We thank Mr. Kenichiro Arakaki from Integrated Open Systems Unit at OIST for providing the real dataset from the houses and Mr. Daisuke Kawamoto from SonyCSL for explaining the APIS simulators.

## REFERENCES

Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

- Vincent François-Lavet, David Taralla, Damien Ernst, and Raphaël Fonteneau. Deep reinforcement learning solutions for energy microgrids management. In *European Workshop on Reinforcement Learning (EWRL 2016)*, 2016.
- ChongAih Hau, Krishnanand Kaippilly Radhakrishnan, JunYen Siu, and Sanjib Kumar Panda. Reinforcement learning based energy management algorithm for energy trading and contingency reserve application in a microgrid. In *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1005–1009. IEEE, 2020.
- Daisuke Kawamoto and Gopinath Rajendiran. A study of battery SoC scheduling using machine learning with renewable sources. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021. URL <https://www.climatechange.ai/papers/icml2021/58>.
- Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, pp. 6379–6390, 2017.
- Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- Taichiro Sakagami, Annette Werth, Mario Tokoro, Yoshiyuki Asai, Daisuke Kawamoto, and Hiroaki Kitano. Performance of a DC-based microgrid system in okinawa. In *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 311–316. IEEE, 2015.
- Taichiro Sakagami, Yoshiyuki Asai, and Hiroaki Kitano. Simulation to optimize a DC microgrid in okinawa. In *2016 IEEE International Conference on Sustainable Energy Technologies (ICSET)*, pp. 214–219. IEEE, 2016.
- Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. *arXiv preprint arXiv:1511.05952*, 2015.
- Sony Computer Science Laboratories, Inc. Autonomous power interchange system. <https://github.com/SonyCSL/APIS>, 2021.
- M Tokoro. SonyCSL-OIST DC-based open energy system (DCOES). In *Proc. 1st Int. Symp. Open Energy Syst*, pp. 64–67, 2014.
- M Tokoro. DCOES : DC-Based bottom-up energy exchange system for community grid. In *2nd International Symposium on Open Energy Systems*, pp. 22–29, 2015.
- Annette Werth, Nobuyuki Kitamura, and Kenji Tanaka. Conceptual study for open energy systems: distributed energy network using interconnected DC nanogrids. *IEEE Transactions on Smart Grid*, 6(4):1621–1630, 2015.
- Annette Werth, Alexis André, Daisuke Kawamoto, Tadashi Morita, Shigeru Tajima, Mario Tokoro, Daiki Yanagidaira, and Kenji Tanaka. Peer-to-peer control system for DC microgrids. *IEEE Transactions on Smart Grid*, 9(4):3667–3675, 2016.
- Suyang Zhou, Zijian Hu, Wei Gu, Meng Jiang, and Xiao-Ping Zhang. Artificial intelligence based smart energy community management: A reinforcement learning approach. *CSEE Journal of Power and Energy Systems*, 5(1):1–10, 2019.