# Deep Reinforced Learning for the Governance of a Sample Microgrid

Berkay Gür, Gulgun Kayakutlu

# Deep Reinforced Learning for the Governance of a Sample Microgrid

Berkay Gür [0000-0002-9524-3640] and Gülgün Kayakutlu [0000-0001-8548-6377]

Istanbul Technical University, Istanbul, Turkey
brky.gur@gmail.com

**Abstract.** A proximal policy optimization reinforcement learning system is proposed to handle the energy dispatch management of a sample microgrid. The microgrid in question has 3 participants of different classifications, signifying their relative importance and how sensitive they are to energy shortages. The energy within the microgrid is generated by these participants, which are individually equipped with a solar panel and a wind turbine for energy generation, and an energy storage system to store this energy. The environmental conditions, i.e. temperature, wind velocity and irradiation figures of Istanbul are considered to obtain accurate energy generation figures. The microgrid is designed to be grid connected in order to compensate for the uncertainties caused by the weather changes, hence service of the utility is accessed when energy produced & stored cannot respond to the demand. Information security of the participants is respected and to that end, direct energy generation, consumption and storage figures are not supplied to the agent, instead only supply and demand figures are transferred. The agent, using this information, after a period of training, optimizes the system for a reward scheme that rewards energy exports and punishes energy deficits and imports. The results verify the feasibility of proximal policy optimization in managing microgrid energy dispatch.

**Keywords:** PPO, Microgrid, Energy Dispatch.

## 1 Introduction

Climate changes caused by Global Warming reduce the quality of life. Residential response to unexpected meteorological pressures has been the quick proliferation of renewable energy sources such as photovoltaic (PV) panels and wind turbines [1]. The energy generation using the renewable sources are, by their very nature, relatively intermittent and, in the case of PV, limited to the daytime. Compared to conventional energy sources, energy generation with renewable energy sources can be achieved at a smaller scale. This enables energy generation in residential areas such as the rooftops. Application of solar on rooftops also has the effect of reducing transmission and distribution losses associated with energy generation, not to mention that they don't take up any further space [2]. Similarly, wind energy can also be utilized in a similar manner [3].

A problem arising with the growth of renewable energy sources is the contradiction with conventional energy generation methods [4]. The traditional utility grid is ill suited to work with the increasing volume of uncoordinated renewable energy generation. One example of this is the phenomenon called the "Duck curve", which refers to the mismatch between the peak demand and renewable energy production. The peak load usually happens at sunset, while renewable energy production reaches its' zenith at mid-evening hours, this causes an energy deficit that can't be satisfied by renewables, requiring the usage of conventional energy sources, and incurring ramp up costs [5].

Hence, microgrids, as a concept, have been designed to cope with the penetration of renewable energy sources [6]. They can be considered as a set of generators, loads and batteries in a single system. This single system can be connected to the larger utility grid at a single point, called a point of common coupling (PCC) and the energy transactions between the utility grid and the microgrid would be managed at that point [7]. It is also possible to manage the microgrid in such a way that enables the microgrid to be independent of the utility grid. This sort of microgrids do not have a PCC since there is no point to be connected to the utility grid. These microgrids are called islanded microgrids. Microgrids with PCC can choose to become islanded by severing the connection between the utility grid and microgrid. This is useful in the cases of catastrophic failure in the utility grid such as massive shortages that may occur due to adverse environmental conditions, such as in the case of Texas power outage of February 2021 [8].

The properties of the grid connected microgrid could be utilized to isolate critical loads for power outages. This paper seeks to use the prioritization property of the microgrid to create a system where the participants of the microgrid are categorized according to their criticality. This would require the microgrid to be constantly supervised for energy deficits and prompt action must be taken to prevent any shortage in the critical loads. To manage this, artificial intelligence could be utilized to automate the monitoring and decision-making aspect of the grid controlling.

To achieve this objective, a solid grasp of microgrids, energy management within the microgrid, utilized methods that exist in the literature should be explored. This requires an organized and methodological survey of the existing literature. Furthermore, this survey would be focused further on the chosen machine learning method, i.e. RL.

## 2 Literature

RL methods are already applied in the field of energy management of microgrids. The energy management within a microgrid, as explained by Murty and Kumar, encompasses both the supply and demand side management of the microgrid [9]. This can, as the authors explain, can involve managing the demand by incentivizing the usage of energy at particular times, or by managing the dispatch, i.e. the flow of the energy that the microgrid controller can utilize. Qazi et. al. [10] proposed a Q-Network system where the supplied energy and the energy present in the energy storage systems are exchanged within an isolated microgrid. Kozlov et. al. [11] simulates an isolated microgrid which contains constant loads, solar panels, wind turbines and biomass engines as a Markov Decision Process automates the dispatch of each energy source. In the

work of Muriithi and Chowdhury the energy in the energy storage system is traded with the prices and the battery degradation taken into consideration [12]. The microgrid system is modeled as a Markov Decision Process and solved using Q-Learning to minimize costs in a successful manner in different tariffs and seasonal conditions.

One of the tasks that can be achieved with reinforcement learning is the management of energy storage systems within the microgrid system. The article authored by Shang et. al. [13] successfully employed Q-Learning with constraints to optimize the load dispatch policies of the energy storage system within a microgrid while taking into account the battery degradation. Research article of Al-Gabalawy [14] applies reinforcement learning along with linear programming to optimize energy transaction decisions of the energy storage systems with both risk averse as well as risk seeking agents. The implementation of energy storage systems are not limited to the use of batteries. Electric vehicles are also explored as in the work of Sadeghi et. al. [15] where a novel Bayesian Coalition Game is explored in microgrids with uncertainties due to EV and attempts to reduce the energy losses within the microgrid. Fan et. al. [16] consider batteries as well as diesel generator within an isolated microgrid with Deep Deterministic Policy Gradient method to minimize the operating cost and compares the result with transfer learning to prove the superiority of the Deep Deterministic Policy Gradient when used in tandem with Transfer Learning.

Samadi et. al. approached this issue from a multi agent perspective [17] where each participant of the microgrid is a seperate agent interacting in a competitive environment. To do this the authors established a Markov Decision Process to find the optimal policy. In this article the authors defined each consumer and energy generator as separate agents and rewards and punishments within the system are the profits and costs of the energy transactions. Fang et. al. similarly employs a multi agent approach for each of the microgrid participants where they interact in a double action scheme for the energy transactions and distributes rewards for consumption and generation from various energy sources. Fang et. al. takes a similar approach with a multi agent system with renewable energy generation participating in energy transactions through an auction scheme [18]. Their work also takes into account the differing load patterns of industry, commerce and residential loads. Guo et. al. [19] modeled the microgrid in a two-level system in a Stackelberg game where the distribution system operator is the leader and the microgrids are the followers and all of the actors in the multi microgrid environment are modeled as a different agent.

Certain design considerations play a huge role in the success of the RL machine learning model. Wu and Wang, in their study, note that the success of a deep reinforcement learning model in the context of microgrids is dependent on the design of the reward structure and badly designed reward schemes could lead to cascading failures [20]. And Yin and Zhang, explain that microgrid operations typically operate on two different time scales, 15-minute time steps for transactions and 4 second steps for generation dispatch, which the authors explain can lead to uncoordinated problems [21]. They attempt to create accurate single time steps with particle swarm optimization and reinforcement learning and compare the results with time series generative adversarial networks. The latter is proposed to generate accurate data that functions on a single time step to improve the operation of the microgrids.

The contributions of this paper to the research on the matter of energy dispatch within a microgrid are twofold.

- First is the lack of information to the grid controlling agent to preserve the personal information of the grid participants.
- Second is the prioritization of the grid members according to their importance to their function or social status, thereby securing the needs according to the specifications of the grid controller.

These contributions do not take into usage of PPO in the microgrid environment, which is, to the best knowledge of the authors, unattempted for the problem of energy dispatching with imports and exports.

The rest of the paper is organized as follows. The second section will be a brief introduction to the topic of reinforcement learning and the used sub-category of reinforcement learning, the proximal policy optimization, along with the reasoning behind the usage of the said methods. The third section will be an examination of the microgrid environment. This section will go over the components of the microgrid, how they are represented to the machine learning algorithm, the observation and action spaces of the reinforcement learning algorithm as well as the reward scheme used in the learning process. The next section consists of a discussion of the results. And the last chapter will be the conclusion and how this model could be further improved in the future.

## 3 Reinforcement Learning

Microgrids, in our case, while using internal energy resources to sustain themselves, remain in connection with the utility grid. This connection allows energy import from the grid at the time of deficit to respond to the demand and export to the grid at the time of excess production. Reinforcement learning is utilized to manage the flow of energy within the microgrids, as well as regulating imported and exported energy.

Reinforcement learning is one the main paradigms of machine learning along with supervised and unsupervised learning [22]. Reinforcement learning is used to carry out trial and error tasks within an environment. How correct are the actions undertaken by the reinforcement learning model is judged by a reward, which itself is a function of the state. The relationship between the chief components of reinforcement learning could be summed up in the Fig. 1.

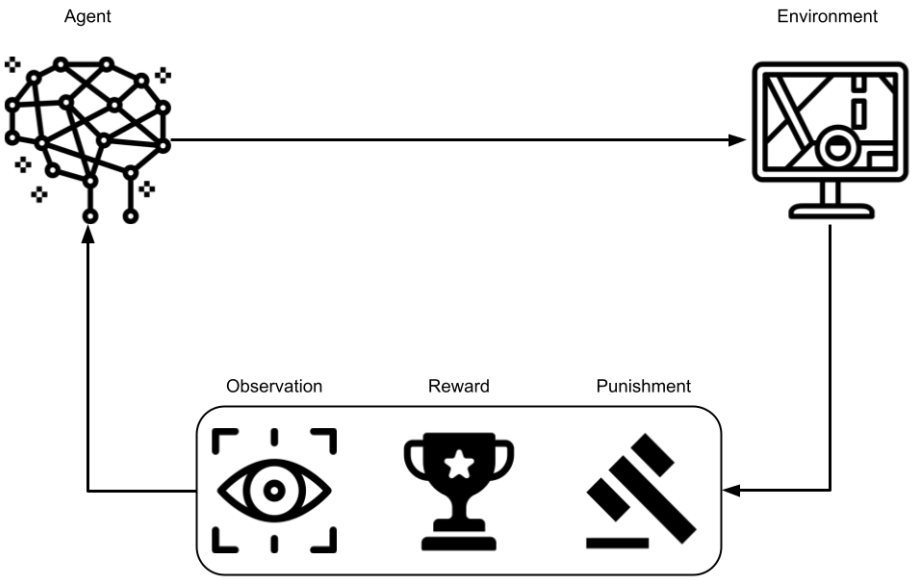


**Fig. 1.** A graphical representation of the reinforcement learning.

In this paper, the proximal policy optimization algorithm is chosen because the learning process itself is a stochastic process. Proximal Policy Optimization is a series of algorithms that build on the previous policy gradient algorithms that are used to learn the policy directly. Rather than dictating the exploration or exploitation through a set of percentages as in the  -greedy method, the policies itself are considered as a probability distribution and each choice made updates the probability, eventually leading to a more stable algorithm than the Q-Learning counterpart [23]. The trade-off is that reaching a suitable solution with policy gradient highly dependent on the step size, if the step size is too low the learning would be very slow and if the step size is too large there may be catastrophic performance implications due to noise [24]. To add on to it, policy gradient algorithms are also sample inefficient, meaning that longer learning time is needed to reach a result compared to the Q-Learning algorithm [25].

Proximal Policy Optimization attempts to solve these issues in policy optimization by utilizing a novel objective function that enables multiple epochs of minibatch updates and reaps the benefits of a trust region policy approach method in a simpler manner with better overall performance [26].

## 4 Microgrid Model

The microgrid, designed for this study, consists of three participants. Each of the participants has its own energy consumption profile, generation, and storage capacities. The participants are also subject to a classification for the criticality.

**Table 1.** Microgrid participant categorization

| Participant Category | Categorization Code |
|---|---|
| Critical | 1 |
| Vulnerable | 2 |
| Normal | 3 |

- The critical participants of the microgrid could be described as the participants that cannot experience any energy shortage under any circumstances. Examples to such microgrid loads are hospitals, server rooms or laboratories that may exist in a university campus. This loss of energy on these loads could put the people within the microgrid in danger, put the scientific progress in an academic setting in jeopardy or potentially disrupt business functions in a commercial setting.
- The vulnerable category is the representation of the participants that potentially can't supply or afford to supply their own energy for heating or for electricity. This definition is similar to the definition established by the EU Energy Poverty Observatory, which states "Energy poverty occurs when a household suffers from a lack of adequate energy services at home." [27]. The lack of energy within these participants would have potential social consequences.
- Lastly, the normal category would represent the participants who can supply or afford to supply their own energy. Although the energy shortage is still undesirable, it can be tolerated for a while and not as critical as the other two users.

All of the participants generate a part of their energy through wind and solar energy. This is calculated on an hourly basis. For solar energy, the Photovoltaic Geographical Information System (PVGIS) is used, which is developed by the European Commission Joint Research Centre and uses the satellite data over Europe and processes the readings into a solar panel power output [28]. To do this, they use data they've collected from solar panels, observed wind velocity, ambient temperature and irradiation. The results of which are logged on an hourly basis for a coordinate over Istanbul, Turkey.

A similar effort is also undertaken from the calculation of wind energy output of the microgrid participants. Energy generated through the wind turbines is dependent on the wind strength of the district. The link between the wind speed and wind turbine output is illustrated as a piecewise function as follows [29].

$$P_{WT,t}\begin{cases} 0 & v_t < v_{ci} \\ P_r \frac{v_t^3 - v_{ci}^3}{v_r^3 - v_{ci}^3} & v_{ci} \leq v_t < v_r \\ P_r & v_r \leq v_t \\ 0 & v_t > v_{co} \end{cases} \tag{1}$$

where $v_{ci}$ and $v_{co}$ are cut in and cut out velocities of the wind turbine respectively. $P_r$ and $v_r$ are rated power and rated velocity of the wind turbine, and $v_t$ is the measured velocity at time of t as processed from the PVGIS dataset for the same coordinate as the solar power output.

The ESS utilized in the system has restrictions in place with regards to the maximum and minimum charge of the energy storage system are implemented as follows.

$$SoC_t \geq SoC_{min} \tag{2}$$

$$SoC_t \leq SoC_{max} \tag{3}$$

The current state of charge $SoC_t$ inside the energy storage system is equal to the summation of the prior state of charge $SoC_{t-1}$ and the incoming or outgoing power $P_{ESS,t}$ multiplied with the efficiency coefficient $\eta$.

$$SoC_t = SoC_{t-1} \pm \eta P_{ESS,t} \tag{4}$$

As is in the case of Kim et. al. [30] 95 percent efficiency for the incoming and outgoing energy is implemented, signified by $\eta$.

Taking all these calculations into account consumer activities are modeled as a workflow. The first step of this workflow is to assess the hourly consumption, solar and wind generation as described previously. The consumers would judge their power balance, if their generated power is enough to cover for the hourly power consumption of the participants. The hourly balance could be formalized as follows.

$$P_{B,t} = P_{C,t} + P_{PV,t} + P_{WT,t} \tag{5}$$

Where $P_{B,t}$ is the power balance in a given time t, $P_{C,t}$ is the power consumption in a given time t, $P_{PV,t}$ and $P_{WT,t}$ is the solar and wind power generation in a given time t respectively.

If the $P_{B,t}$ is negative, meaning there is an energy deficit between the generated power and consumed power at the given time t, then the deficit will be compensated from the energy storage system. Next, the participant in the deficit will decide whether the energy in the energy storage system is sufficient. If the energy storage systems projected state of charge goes below the minimum level, the energy storage will reach minimum, and the remaining deficit will be satisfied by the microgrid.

Conversely, if the $P_{B,t}$ is positive, that is to say there is a surplus of power in the participant, the participant will initially use this excess to charge its energy storage system. If the energy storage system is full, then the excess will be forwarded to the microgrid for usage or export.

Therein lies one of the novelties of the paper, the reinforcement learning environment does not convey the information of the energy consumption and generation to the agent. The grid manager effectively has no knowledge about the energy consumption or generation capabilities of the microgrid participants, providing a layer of privacy to the participants.

In summation, the way those participants of the microgrid function in the model can be illustrated as a flowchart in Fig. 2.

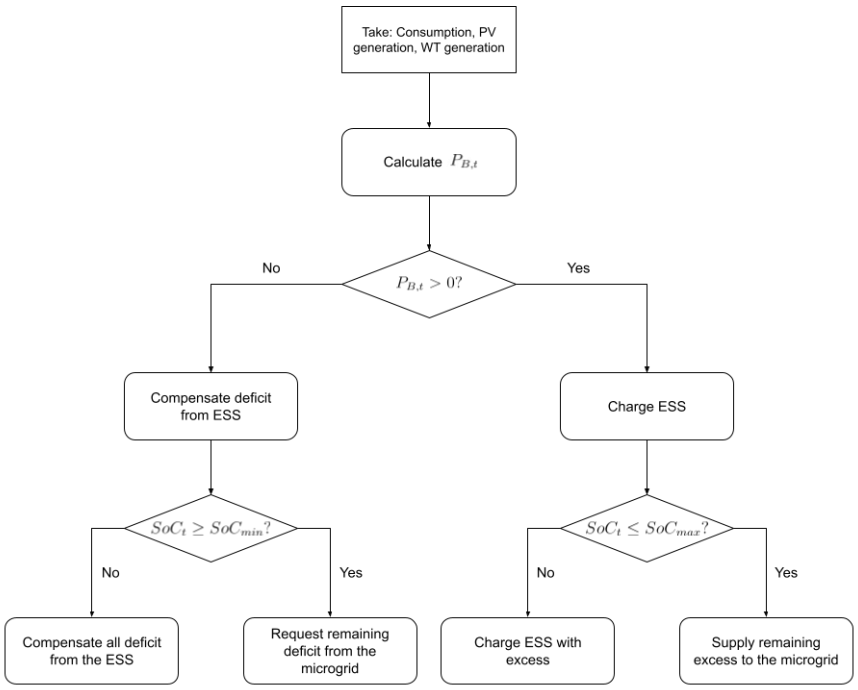


**Fig. 2.** Flowchart of microgrid consumer activities.

After modelling the microgrid, modelling of the agent and its action and observation space must be formally defined.

As we've briefly touched on previously, privacy is integrated into the design of the machine learning algorithm. As a result the only data the agent sees from the environment, that is to say from the microgrid, is the power requests, supply offers and the category descriptions of the individual participants.

Power requests and offers are collected individually for each participant. Considering the sample microgrid consists of 3 participants, the observation space would be a 1x3 matrix. Since the energy balances of the participants are unbounded, this observation space can be summarized thusly where the $B$ is the power balance at any given time t.

$$[B_1, B_2, B_3] \quad \forall \quad B \in \mathbb{R} \tag{6}$$

The agent does not take into account just the energy balance of any given hour, but also the categorization of the individual participants of the grid. Observation space for it is as follows. The $C$ in this case represents the category of the participant.

$$[C_1, C_2, C_3] \quad \forall \quad C \in [1, 2, 3] \tag{7}$$

where 1,2,3 are the same codes for the participant categorizations in table X. Taking all of these into account the observation space that the agent observes is a 1x6 matrix with the previously mentioned characters.

Once the observation space is established, the agent will need to act according to the knowledge it receives from the observation space. The actions the agent can take in this context can be divided into two.

First action the agent can take is to divide the internally offered energy to the microgrid participants that may need the energy, thereby using the excess internally. This action is represented as a 1x4 matrix. The first three columns are for the amount of power the participants receive from the internal surplus. The fourth column is for the total energy exported outside of the grid for reward. This reward is assigned to teach the agent that any excess energy exported outside the grid will have a monetary benefit to the microgrid, hence it's being rewarded. This matrix could be summed up as follows and $I$ is shorthand for internally used power and $X$ is shorthand for exported power.

$$[I_1, I_2, I_3, X] \quad \forall \quad I, X \in \mathbb{R} \tag{8}$$

For the cases where the internal surplus energy is not enough, externally supplied energy is needed, for these cases the agent can assign how much power each of the participants will have from outside the grid at each time step t. The imported power $M$ is added to the action space as a 1x3 matrix thusly,

$$[M_1, M_2, M_3] \quad \forall \quad M \in \mathbb{R} \tag{9}$$

Once the appropriate power is supplied to the grid through the action space, the actions will be judged by the environment and rewards will be assigned.

A well-designed reward system, as was explained in [20], is of utmost importance for the functioning of the machine learning model. The rewards and punishments are used to guide the agent to a policy, that is to say internal to logic, that would be in line with how a microgrid should be run. To do this, first rewards should be assigned to successful usage of internal excess energy within the grid. If there is still excess energy within the microgrid, then the rest should be exported. Lastly, in the case of energy shortages within the microgrid, external sources should be utilized to balance the deficit. To achieve this goal the following rewards are assigned.

**Table 2.** Rewards for the actions.

| Participant Category | Categorization Code |
|---|---|
| Critical | 1 |
| Vulnerable | 2 |
| Normal | 3 |

This reward scheme alone is not enough to support the activity of the microgrid, as there is no punishment for simply not supplying the participants in case of an energy deficit within the microgrid. To make sure the agent will import energy from outside despite the negative reward, i.e. punishment, specific negative rewards are assigned to each category of microgrid participants. This auxiliary reward scheme could be summed up as follows.

**Table 3.** Negative rewards for power shortages in any given time for one hour.

| Participant Category | Categorization Code |
|---|---|
| Critical | 1 |
| Vulnerable | 2 |
| Normal | 3 |

With these configurations, the agent is set to explore and learn from the environment to manage the microgrid.

# 5 Results

After taking 15 million actions and receiving rewards from the environment, the agent is introduced to a different data set of consumer data. The process of learning with 15 million cases takes around a day while returning with a course of action for the test data is almost instantaneous. The results are recorded for 48 consecutive hours for one sample episode. The results for the entire microgrid could be observed in figure 3.

The first thing to note in the figure is that the model prefers to use a mixture of internal and externally sourced energy for most of the episode. This may be due to a few reasons. The most likely explanation for this behavior could be that the model tries to avoid not being able to supply an adequate amount of energy to the participant. To achieve this goal the model purposefully takes a less than optimal route to make sure it does not incur the penalty of not being able to supply sufficient amounts of energy to the system.

Similarly, after the tenth hour of the episode, we begin to observe energy exports. It is possible that the model has recognized the daily pattern of energy generation and consumption and decided that it is the optimal time slot to begin exporting the excess energy that it possesses in order to maximize the amount of reward earned in the episode. It is worthwhile to note that the model continued importing energy into the microgrid while exporting at the same time. It is possible that the agent has chosen this course of action due to the fact that there is more reward in exporting energy than there is punishment for importing, thereby it seems logical to the model to export the energy at hand and import a token amount to make sure that there is no energy deficit within the grid.

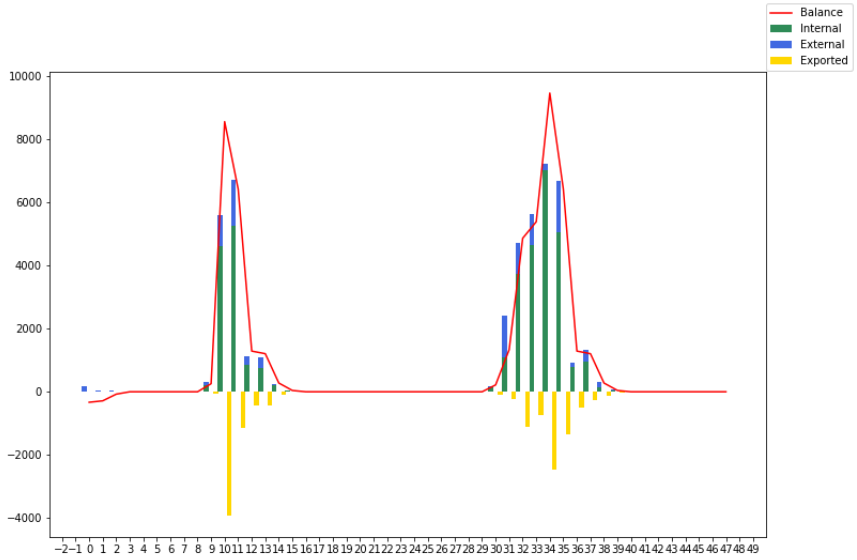

**Fig. 3.** Overall energy balance of the microgrid.

A similar picture also exists for the critical participant of the microgrid, a great excess is consistently maintained by the grid controller, presumably to make sure there is no shortage in the system. There are moments where not all of the supply is used neither by the internal nor by the external consumption, causing a space between line and bars. This is due to the fact that excess energy is exported to maximize the reward of the microgrid. Despite the export of energy, we also observe sharing other participants' production, shown as green bars. Adding to that some amount of energy is imported from outside the microgrid to make sure there is no shortage in the microgrid.

Furthermore, throughout the first day there is no supply and demand, pointing to the fact that excesses in that time are used to charge the energy storage systems within the grid. The next morning the excesses left after the energy storage systems are fully charged some amount of energy is coming from the grid controller and the excess is exported.

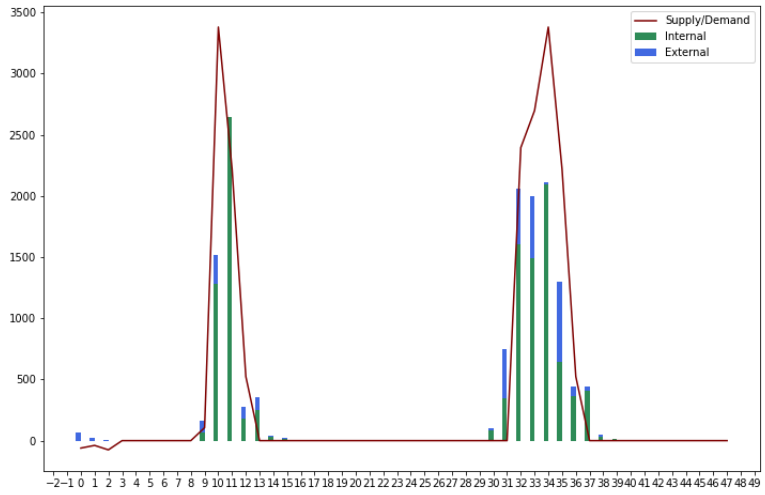


**Fig. 4.** Overall energy balance in the critical participant of the microgrid.

A similar picture is also seen with the vulnerable participant of the microgrid. In graph 5 it is observed that a greater liberty is taken with the excess amount, as seen by the larger gap between the supply line and the supplied amount seen as the bar. The difference in between is sold as export to the grid.

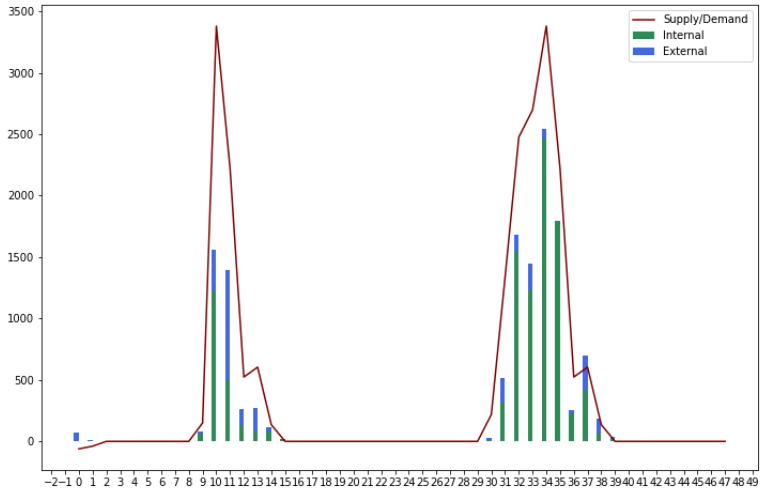


**Fig. 5.** Overall energy balance in the vulnerable participant of the microgrid.

Mismatches can be observed in the activities when it comes to the normal participant of the microgrid. While the agent is mostly successful in making sure there is no deficit in the grid, there exists cases where the supply and demand are mismatching, as can be seen in graph 6.

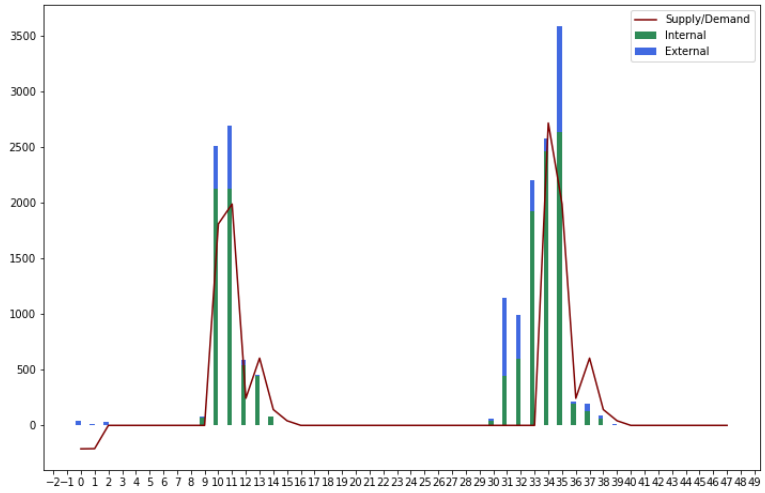


**Fig. 6.** Overall energy balance in the normal participant of the microgrid.

## 6 Conclusions

Climate change caused by global warming is an ever-increasing problem in the global community. One method at our disposal in overcoming this problem is the increased usage of renewable energy sources in the energy mix. To spare the utility grid of the complications of the more intermittent nature of renewable energy generation, this paper takes into account a sample microgrid where these energy sources would be utilized. The model proposed in this paper attempts to automate the operations within this microgrid with regards to energy dispatch, as well as the energy imports and exports.

To that end, reinforcement learning is chosen due to its reward-based structure and ability to make decisions based on incomplete knowledge. Reinforcement learning algorithm chosen for this application is the proximal policy optimization. The reason for this choice is the fact that it's newer than the other algorithms that have been covered in this paper, such as Q-Learning. The PPO also has the distinction of being a policy optimization algorithm which means that the model, rather than maximizing directly for the reward earned from each move, optimizes itself for the probability of earning the most reward at the end of the episode, without falling to the common pitfalls faced by other policy optimization algorithms such as low sample efficiency and performance issues.

A simulation of the microgrid system is built specifically for the sample microgrid using OpenAI Gym. Here lies one of the novelties of this paper, the RL learning environment constructed for this application does not convey the information of energy generation and consumption directly to the agent. The only information received by the

agent is the supply and demand of the participants and their corresponding criticality. This criticality information depends on the nature of the participant and ranges from critical to normal depending on whether any energy losses are acceptable considering their importance to the grid, and to the fact whether they are at risk of not being able to supply their own energy. Such categorization, to the best of the author's knowledge, does not exist in other RL applications in the field of microgrid energy dispatch.

The agent is set to perform 15M timesteps in different environmental conditions with a variety of energy consumption profiles to learn how to manage the energy dispatch within the microgrid. The performance of the system displays that the proximal policy optimization algorithms can feasibly be used in the governance of a microgrid. While it is possible to observe cases where participants were importing energy while there is an energy surplus, this can potentially be attributed to the uncertainty of the environment.

This is the first work to feature PPO and participant criticality within the microgrid. This work could be further in the future with the improvement reward system to include stricter punishments for the oversupply of the energy, implementation of a carbon tax to further prioritize renewable energy sources and the usage of publicly owned ESS and renewable energy sources, which could be used either to extract revenue from the system, or to further reduce the energy dependency of the microgrid.

## References


1. A. Sahin: Progress and recent trends in wind energy. Progress in Energy and Combustion Science, vol. 30, no. 5, pp. 501–543 (2004).
2. A. C. Duman and Ö. Güler: Economic analysis of grid-connected residential rooftop PV systems in Turkey. Renewable Energy, vol. 148, pp. 697–711 (2020).
3. A. Rezaeiha, H. Montazeri, and B. Blocken: A framework for preliminary large-scale urban wind energy potential assessment: Roof-mounted wind turbines. Energy Conversion and Management, vol. 214, p. 112770 (2020).
4. C. P. Vineetha and C. A. Babu: Smart grid challenges, issues and solutions. In 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG), Taipei, Taiwan, pp. 1–4 (2014).
5. A. Majzoobi and A. Khodaei: Application of Microgrids in Supporting Distribution Grid Flexibility. IEEE Trans. Power Syst., vol. 32, no. 5, pp. 3660–3669 (2017).
6. J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuna, and M. Castilla: Hierarchical Control of Droop-Controlled AC and DC Microgrids—A General Approach Toward Standardization. IEEE Trans. Ind. Electron., vol. 58, no. 1, pp. 158–172 (2011).
7. Farrokhabadi M, Canizares CA, Simpson-Porco JW, Nasr E, Fan L, Mendoza-Araya PA, et al.: Microgrid Stability Definitions, Analysis, and Examples. IEEE Trans. Power Syst., vol. 35, no. 1, pp. 13–29 (2020).
8. Wu D, Zheng X, Xu Y, Olsen D, Xia B, Singh C, et al.: An open-source extendable model and corrective measure assessment of the 2021 texas power outage. Advances in Applied Energy, vol. 4, p. 100056 (2021).
9. V. V. S. N. Murty and A. Kumar: Multi-objective energy management in microgrids with hybrid energy sources and battery energy storage systems. Prot Control Mod Power Syst, vol. 5, no. 1, p. 2 (2020).

10. H. S. Qazi, N. Liu, and T. Wang: Coordinated Energy and Reserve Sharing of Isolated Microgrid Cluster using Deep Reinforcement Learning. In 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, China, pp. 81–86 (2020).
11. A. N. Kozlov, N. V. Tomin, D. N. Sidorov, E. E. S. Lora, and V. G. Kurbatsky: Optimal Operation Control of PV-Biomass Gasifier-Diesel-Hybrid Systems Using Reinforcement Learning Techniques. Energies, vol. 13, no. 10, p. 2632 (2020).
12. G. Muriithi and S. Chowdhury: Optimal Energy Management of a Grid-Tied Solar PV-Battery Microgrid: A Reinforcement Learning Approach. Energies, vol. 14, no. 9, p. 2700 (2021).
13. Y. Shang et al.: Stochastic dispatch of energy storage in microgrids: An augmented reinforcement learning approach. Applied Energy, vol. 261, p. 114423 (2020).
14. M. Al-Gabalawy: Advanced machine learning tools based on energy management and economic performance analysis of a microgrid connected to the utility grid. Int J Energy Res, p. er.6764 (2021).
15. M. Sadeghi, S. Mollahasani, and M. Erol-Kantarci: Power Loss-Aware Transactive Microgrid Coalitions under Uncertainty, Energies, vol. 13, no. 21, p. 5782 (2020).
16. L. Fan, J. Zhang, Y. He, Y. Liu, T. Hu, and H. Zhang: Optimal Scheduling of Microgrid Based on Deep Deterministic Policy Gradient and Transfer Learning. Energies, vol. 14, no. 3, p. 584 (2021).
17. E. Samadi, A. Badri, and R. Ebrahimpour: Decentralized multi-agent based energy management of microgrid using reinforcement learning. International Journal of Electrical Power & Energy Systems, vol. 122, p. 106211 (2020).
18. X. Fang, Q. Zhao, J. Wang, Y. Han, and Y. Li: Multi-agent Deep Reinforcement Learning for Distributed Energy Management and Strategy Optimization of Microgrid Market. Sustainable Cities and Society, vol. 74, p. 103163 (2021).
19. C. Guo, X. Wang, Y. Zheng, and F. Zhang: Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning. International Journal of Electrical Power & Energy Systems, vol. 131, p. 107048 (2021).
20. T. Wu and J. Wang: Artificial intelligence for operation and control: The case of microgrids. The Electricity Journal, vol. 34, no. 1, p. 106890 (2021).
21. L. Yin and B. Zhang: Time series generative adversarial network controller for long-term smart generation control of microgrids. Applied Energy, vol. 281, p. 116069 (2021).
22. R. S. Sutton and A. G. Barto: Reinforcement learning: an introduction. 2nd Edition. MIT Press, Cambridge (2018).
23. S. Gu, T. Lillicrap, Z. Ghahramani, R. E. Turner, and S. Levine: Q-Prop: Sample-Efficient Policy Gradient with An Off-Policy Critic. ArXiv (2017).
24. Open AI Proximal Policy Optimization Webpage, https://openai.com/blog/openai-baselines-ppo, last accessed 02/09/2021.
25. Introduction to Deep Reinforcement Learning (Deep RL), https://www.youtube.com/watch?v=zR11FLZ-O9M&t=3487s, last accessed 02/09/2021.
26. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov: Proximal Policy Optimization Algorithms. arXiv (2017).
27. EU Energy Poverty Observatory, What is energy poverty?, https://www.energypoverty.eu/about/what-energy-poverty, last accessed 01/09/2021
28. Photovoltaic Geographical Information System (PVGIS), https://ec.europa.eu/jrc/en/pvgis last accessed 02/09/2021.
29. R. Atia and N. Yamada: Sizing and Analysis of Renewable Energy and Battery Systems in Residential Microgrids. IEEE Trans. Smart Grid, vol. 7, no. 3, pp. 1204–1213 (2016).

30. R.-K. Kim, M. B. Glick, K. R. Olson, and Y.-S. Kim: MILP-PSO Combined Optimization Algorithm for an Islanded Microgrid Scheduling with Detailed Battery ESS Efficiency Model and Policy Considerations. Energies, vol. 13, no. 8, p. 1898 (2020).