Trading Strategy in a Local Energy Market, a Deep Reinforcement Learning Approach

Olamide Jogunola*, Yakubu Tsado*, Bamidele Adebisi*, Raheel Nawaz**
*Department of Engineering, Manchester Metropolitan University, Manchester, UK.
Email: {o.jogunola, y.tsado, b.adebisi}@mmu.ac.uk
**Business School, Manchester Metropolitan University, Manchester, UK.
Email: r.nawaz@mmu.ac.uk

***Abstract*—In response to energy transition fueled by the increasing energy generation mix and dynamic environment, this paper presents an energy trading strategy utilising real microgrid data. Specifically, we adapted the deep Q-network (DQN) with prioritised experience replay (PER) to develop a DQN-PER-based energy market algorithm to optimise the utility derived by prosumers participating in a local energy market (LEM). The problem of exercising energy trading actions is formulated as a sequential decision-making problem to optimise the prosumer's utility in a variety of energy trading scenarios. This includes the contingency or flexibility provided by the energy storage system (ESS), the incorporation of solar photovoltaic (PV) sources and the decision to trade energy with the grid or in a LEM. The results show the benefit achieved in trading energy in LEM with higher benefits when more sources of renewable energy are incorporated. For instance, the average benefit of trading in the LEM over the grid with ESS is 35%, which increased to 54% when PV and ESS are incorporated.**

***Index Terms*—Solar PV, energy storage system, deep-Q network, local energy market, deep reinforcement learning, priortised experience replay.**

I. Introduction

The energy trilemma faced by power networks and their customers involves: 1) transitioning to zero carbon emission; 2) lowering energy cost; and 3) providing secure and reliable supply to meet growing demand [1]. These factors are beginning to drive behavioural change, particularly among energy consumers transitioning into prosumers. A producer and consumer of energy with choice to control its energy consumption, generation, emissions and revenue from energy sale [2]. Changes to consumption and energy control pattern increase power sector complexity. Resulting in a need for flexibility and integration of new tools for system optimisation and energy trading [3]. Addressing these changes necessitate real-time control and supervision in energy grid management and operation at medium and low voltage levels. A constrained optimisation approach is mostly proposed for optimal control of power sector complexity. These constrained optimisation include linear or nonlinear mathematical programming, mixed integer programming or dynamic programming, with set objectives to minimise cost [4], improve efficiency [5], optimal energy schedule [6] or energy trading [7]. While these models solve the control complexity, they relatively require a high level of mathematical and system-wide knowledge, which might not be applicable in a fast, dynamic distributed energy trading environment [8].

Meanwhile, the recent advancement in artificial intelligence, mainly reinforcement learning (RL) has emerged for (near) optimal control of dynamic systems in energy network including energy trading [9], [10]. This is achieved by transforming the energy control or trading problem into a sequential decision-making problem and using RL to solve it. Here, prosumers can exercise trading actions without extensive knowledge of the market system model, nor analytical optimisations. The Authors in [11] proposed an indirect customer-to-customer energy trading in a localised event-driven market, which is solved using RL to maximise the customer's benefit. RL has been used extensively in other energy network applications including management and control, but not extensively in the energy market. This is a result of its ineffectiveness in high-dimensional state-action transitions. RL computational time increases as the network grows. Thus, a neural network is usually adopted to extend the RL to deep-RL (DRL) in the energy market. For instance, the authors of [12] presented a prosumer trading behaviour in a local energy market utilising DRL, where features of DRL are explored in a data-driven market model. Likewise, [13] proposed a DRL based on a deep Q-network (DQN) peer-to-peer (P2P) energy trading model for microgrid decision making and [14] applied a DQN for agents in a consumer-centric energy market.

While these studies adapted DRL in energy market trading behaviours, the contingency provision of energy storage system (ESS) with PV generation and dynamic pricing, resulting in multi-dimensional state-action spaces require further impact investigation on the annual benefits of their acquisitions. Besides, these studies utilised experience replay during the DQN training which only stores the state, action and reward of training. As opposed to previous DRL-based energy trading schemes, we employed prioritised experience replay (PER) that places priority on the experiences with high errors which have been shown to improve and accelerate DQN training in

This work was supported in part by ENERGY-IQ, a UK-Canada Power Forward Smart Grid Demonstrator project funded by The Department for Business, Energy and Industrial Strategy (BEIS) under Grant number:7454460, and in part by the NICE (Nigerian Intelligent Clean Energy) Marketplace project funded by the Department for International Development (DFID).

other cyber-physical applications [15]. The contributions of this work are as follows:

- We propose a DRL-based algorithm for the local energy market (LEM) to optimise the utility derived by trading energy locally. In particular, we proposed a DQN-PER energy trading as opposed to experience replay used in studies [11]–[13]. This prioritisation improves and accelerates the DQN training thereby reducing the computation time.
- By investigating ESS as a contingency provision, we design a sequential decision-making problem to achieve energy trading strategy in response to dynamic pricing and the available resources while studying its effect on the annual net benefit derived by the prosumers.
- The proposed DQN-PER-based energy trading algorithm is evaluated using the PECAN[1] microgrid data [16] for a variety of scenarios. The proposed model provides optimal trading action based on prosumer ESS, generation and consumption information for each designated trading time unit. The result is analysed for energy trading with the grid and trading in the LEM.

The remaining sections are organised as follows. A description of the prosumer model is presented in Section II. The proposed DQN-PER-based energy trading algorithm is presented in Section III followed by the discussion of the simulation and result in Section IV. Section V concludes the paper with future work.

II. Prosumer model

The local energy market design represents a pool of local energy balancing operated at the distribution level by the system operator. We consider N prosumers and the utility grid in the LEM, connected through the transmission network. Each prosumer is equipped with an ESS, a PV system and flexible load. The prosumers are responsible for the charging and discharging actions of their ESS to improve their utility during self-consumption or energy trading. Based on their flexibility and modelling, prosumers can strategically decide to trade energy in the LEM or with the grid. The following subsections model the ESS and defined the problem formulation.

A. Energy Storage System Model

Flexibility provision is mostly aided by an ESS, which forms a core part of energy contingency provision. This is constraint by (1), i.e. maximum, and minimum charging limit of the storage capacity. This charging constraint significantly impacts the performance of the trading objectives.

$$SoC_{min} \leq SoC \leq SoC_{max} \tag{1}$$

The ESS model follows (2) [12] for SoC at time t:

$$SoC_{t+\Delta t} = SoC_t + \frac{E_{ch}\rho_{ch}\Delta_t}{CAP_{bat}} - \frac{E_{dis}\Delta_t}{CAP_{bat}\rho_{dis}} \tag{2}$$

[1]PECAN dataset is for residential households that covers a variety of load patterns used at variable times of the day

where E_{ch} and E_{dis} are charging and discharging power respectively. CAP_{bat} is the capacity of the battery, ρ_{ch} and ρ_{dis} are the charging and discharging efficiency respectively. In addition, the empirical wear-cost coefficient k of the ESS will be taking into account in the energy trading decisions. This is expressed as:

$$k = \frac{C_e}{\rho_{dis} CAP_{bat} L_i \delta} \tag{3}$$

where C_e is the ESS capital cost (£/kWh), L_i is the number of the life cycles, and δ is the depth of charge of the ESS.

B. Utility function and energy trading actions

A prosumer can choose to buy or sell energy with the grid or in the LEM with the consideration of the surplus or deficit of the energy sources including ESS, PV and demand requirement. The four energy trading actions are discussed in Table I.

TABLE I
Energy trading actions

Action	Objective
Action 1	To decide when to charge/ discharge the ESS from/to the grid
Action 2	To decide when to integrate the PV with Action 1
Action 3	To decide when to charge/ discharge the ESS from/to the LEM
Action 4	To decide when to integrate the PV with Action 3

The utility function u for Actions 1 and 2 is expressed in (4) where Pg_t is the grid price, Pm_t is the price from LEM and Z is a preset multiplier of probable action at a time t. Also, PV_t is the energy generated from the PV at time t and ESS_t is the current SoC of the battery at time t. k is as defined in (3).

$$\begin{aligned} &u(Pg_t Pm_t, ESS_t, PV_t, D_t|A_t) = \\ &(Pg_t - Pm_t) \times Z - k \times (CAP_{bat} - SoC) \end{aligned} \tag{4}$$

Z is further expressed in (5).

$$Z = \begin{cases} \dfrac{(CAP_{bat} - SoC)}{\rho} + D_t & \text{for Actions 1,3 (5a)} \\ \dfrac{(CAP_{bat} - SoC)}{\rho} + PV_t\Delta_t - D_t & \text{for Actions 2,4 (5b)} \end{cases}$$

Where ρ is the battery efficiency, D_t is the total demand at time t and $PV_t\Delta_t$ is the change in PV levels at time t. Similar to [12], (5b) represent Action 1, where the ESS is mainly charged/discharged from/to the grid as there is no form of onsite generation to satisfy the demand. In Action 2, there is an onsite generation, however, considering the wear cost of the ESS, it might be more economical to charge from the grid when it is cheaper especially when the onsite generation is not sufficient to satisfy the demand. The utility for Actions 3 and 4 is expressed as:

$$\begin{aligned} &u(Pg_t Pm_t, ESS_t, PV_t, D_t|A_t) = \\ &(Pm_t - Pg_t) \times Z - k \times (CAP_{bat} - SoC) \end{aligned} \tag{6}$$

where Z is expressed for Actions 3 and 4 respectively in (5). For Actions 3 and 4, the main trading benefit is a result of the reduced price in buying from the LEM against the grid.

C. Net Benefit of a Prosumer

The ultimate aim of the prosumer is to utilise its available resources to maximise the total utility or economic benefits of owning a PV or ESS achieved either through self-consumption or local energy trading. The utility maximisation problem can be expressed as:

$$\textbf{max} \quad U_t = \sum_{t=0}^{T} u(Pg_t Pm_t, ESS_t, PV_t, D_t| \\ \pi(Pg_t Pm_t, ESS_t, PV_t, D_t)) \tag{7}$$

Utilising the historical data of the PV and load assets, the market price and the battery model, the DRL agent can be trained to respond to future uncertainties and strategic decisions on future energy usage and trading.

III. DQN-PER-BASED ENERGY TRADING STRATEGY

DRL utilises the data collected from real-world cyber-physical systems to model and solve stochastic control problems with uncertainties. DRL model is a data-analytic framework that usually follows a Markov Decision Process (MDP), satisfying Markov's property which highlights the relationship between the previous, the current and the future states [17]. The proposed DQN-PER energy trading model integrated into the DRL environment summarised in Fig. 1 is discussed in the following subsections.

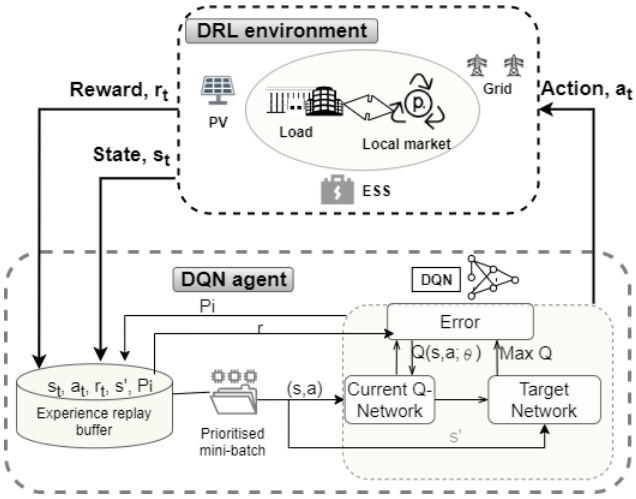

Fig. 1. Proposed DQN-PER-based energy trading.

A. Proposed DQN-PER-based energy trading model

MDP, defined as a tuple of (S, P, R, A, γ) is an environment where the agent interacts. S is the current state of the agent, P is the probability matrix of transitioning from a state S to next state S', R is the reward function of performing an action A and γ is the discount factor of accepting an immediate or a later reward. The agent in the MDP environment, like an energy trading network, in this case, utilises different strategies or policies to maximise its total rewards. The total reward called Q-value is expressed as the Bellman equation (8)

$$Q(s,a) = R(s,a) + \gamma \max_a Q(s',a) \tag{8}$$

where s' is the state transition to the next state. Equation (8) presents the Q-value of state-action space as the immediate reward $R(s,a)$ and the maximum Q-value from s'. Equation (8) can be further expressed as the Bellman iterative equation as

$$Q(s,a) = Q(s,a) + \alpha[R(s,a) + \gamma \max_a Q(s',a') - Q(s,a)] \tag{9}$$

where α is the learning rate or step size. To enhance the training of the Q-learning and to approximate the Q-value function in DRL, a deep neural network (DNN) is utilised. This makes the network more stable and data-efficient, by utilising experience replay and a separate target network. Here, we train the Q-network by minimising a sequence of loss functions $L_t\omega_t$ over the iteration t. Using (9), the error function is calculated as the difference between the maximum possible value from the next state (Q_target) and the Q_value (current prediction) expressed in (10)

$$L_t(\omega_t) = E_{s,a,s',r\sim D}\left[\theta_t - Q(s,a;\omega_t)\right]^2 \tag{10}$$

where $\theta_t = r + \gamma \max_a Q(s',a;\omega_{t-1})$ is the target for iteration t. Equation (10) samples the environment and stores the observed experiences in a replay memory, then a small batch is selected for learning using a gradient descent update step.

In contrast to the experience replay mostly adopted in the literature, a prioritised experience replay (PER) is used in this study. In addition to the state, action and reward variables stored by the experience replay, PER also stores the loss function/error of each experience as shown in Fig. 1, while placing priority on those experiences with high error. To determine the priority P_i of each experience being selected, a minimal constant error er is added to the magnitude of error $L_t(\omega_t)$ [18]. This is expressed in (11).

$$P_i = |L_t(\omega_t)| + er \tag{11}$$

To prevent overfitting and to ensure the probability Pr_i of the selected experience considers the priority and the randomness hyperparameter ρ, a stochastic prioritisation is adapted to replace the greedy prioritisation as follows:

$$Pr_i = \frac{P_i^\rho}{\sum P_i^\rho} \tag{12}$$

Where ρ is a hyperparameter used to reintroduce some randomness in the selection of experiences for the replay buffer. The highest priority experience is selected when $\rho = 1$ and a random selection when $\rho = 0$. However, the DQN can be overfitted due to the bias introduced by the priority sampling. To correct the bias, the weights of the experience ω_t is adjusted as follows:

$$\omega_i = \left(\frac{1}{B_s \times Pr_i}\right)^\beta \tag{13}$$

Where B_s is the batch-size and β is the sampling control rate of each learning.

Algorithm 1: The DQN-PER algorithm for local energy market.

Input: minibatch m, step-size, replay period T, exponents ρ and β.
Initialise a tree-based trading replay memory D to capacity N
Calculate SoC of ESS using (2) and (3)
Observe $S_0 = \{Pg_0Pm_0, ESS_0, PV_0, D_0\}$ and choose a random A_0 from (5)
for $t = 1, \cdots, T$ **do**
    Observe next state, S_t, γ_t and reward, R_t using (4) and (6)
    Store transition $(S_0, A_0, R_t, \gamma_t, S_t)$
    for $j = 1, \cdots, m$ **do**
        Sample transition with probability Pr_i in (12)
        Compute importance-sampling weight using (13)
        Compute absolute loss of (10)
        Update transition priority using (11)
        Accumulate weight change
    end
    Calculate annual benefit based on A_t using (16) and (17)
end

B. Net Benefit of a Prosumer as an MDP

Formulating the energy trading network as an MDP, the state S consists of a time component, load component, generation component, SoC of battery, and the energy cost [3], [17]. The state space is defined as:

$$S = S^t \times S^b \times S^p \times S^c \times S^l \tag{14}$$

where S^t is the daily time component that provides the information to learn the energy generation and consumption pattern, S^b is the SoC of the ESS component, S^p is the PV energy generation component, S^c is the cost component that determines the cost of electricity (buy/sell), and S^l is the energy demand/load component. The action space of the MDP is the energy trading decisions. The trading action space includes buying/selling from/to LEM or the grid $U = \{-1, -0.5, -0.25, 0.25, 0.5, 1\}$, where -1 and -0.5 are the actions to buy/sell in LEM, 1 and 0.5 are the action to buy/sell to the grid, ± 0.25 is the action to incorporate the PV as defined in Table I. The reward function R is the incentive provided when some actions are performed. For instance, reduction in cost by charging the battery using onsite generation, or cost saved when using a battery storage device rather than grid consumption.

Thus, representing problem (7) as an MDP can be written as:

$$\textbf{max} \quad U_t = \sum_{t=0}^{T} \gamma^t r_t(s_t, \pi(s_t)) \tag{15}$$

where π is the policy that maps state to an action; $\pi : S \longrightarrow A$, and state $s_t = (Pg_tPm_t, ESS_t, PV_t, D_t)$. The annual net benefit η of performing Actions 1 and 2 is expressed in (16)

$$\eta(s,a) = p_{sell}E_{ToGrid}(s,a) - p_{buy}E_{FromGrid}(s,a) \tag{16}$$

where E_{ToGrid} is the energy injected to the grid, and $E_{FromGrid}$ is the energy from the grid. While for Actions 3 and 4 is expressed in (17)

$$\eta(s,a) = p_{sell}E_{ToLem}(s,a) - p_{buy}E_{FromLem}(s,a) \tag{17}$$

where E_{ToLem} is the energy injected to the LEM, and $E_{FromLem}$ is the energy from the LEM.

The full proposed DQN-PER energy trading model is summarised in Algorithm 1.

IV. Results and Discussion

In this section, the dataset utilised for the study is first discussed followed by the simulation results of the proposed model.

A. Data Description and System Parameters

An excerpt dataset for a prosumer used to evaluate the model is presented in Fig. 2, showing the PV generation, building demand and price of electricity for the prosumer in a 24hr period. While the consumption and PV generation data

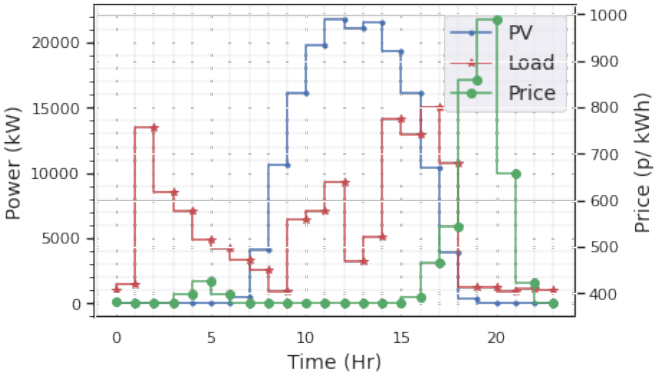

Fig. 2. 24hr data sample, showing the PV, load and price for house 7.

for prosumers are from the PECAN microgrid [16], the system parameters for the ESS are presented in Table II.

TABLE II
System Parameters for the ESS

Capacity	Efficiency	Lifetime
80-100kWh	90.6%	10 years

The computation to train and test the DQN-based flexibility model is performed on Google Colaboratory [19] using Intel Core i7-CPU, 16 GB RAM and 64-bit operating system. After extensive experimentation, the following parameters are used. The DQN neural network contains 1 input layer with 5 neurons taking up the state space values consisting of the grid price, PV, load, ESS and the LEM price. There are two fully

connected hidden layers with 40 and 80 neurons respectively. 1 output layer for the predicted Q-value using 11 neurons. ReLU activation function for the first three layers and linear activation function for the output layer. Other hyperparameters include α, 0.001; B_s, 64; β, 0.4; γ, 0.95; and ρ, 0.6.

B. Optimal trading strategy

Utilising the PECAN microgrid data [16] over a year for 10 houses and assuming the LEM price offer is 70% of the grid prices, we optimise and analyse the economic benefit for each prosumer if energy is traded with the grid or in the LEM. Fig. 3 shows the load profile for the first 5 houses over a 24 hr period illustrating variance in the household energy consumption.

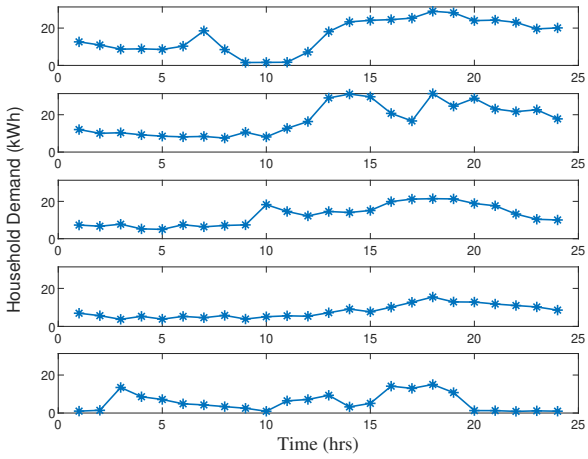

Fig. 3. 24hr load demand for first five houses showing variable demands.

With the proposed DQN-PER trading Algorithm 1, prosumers can exercise energy trading decisions without explicit knowledge of the market model or analytical calculations. This is because DRL utilises historical experience in training the model thereby bypassing the complexity of model-based optimisation. In Algorithm 1, the replay memory is set to a trading environment where the action is to utilise the historical prices of the grid and the LEM to decide on the best strategy to maximise utility. For instance, based on the price offerings and the available generation and demand conditions, the prosumers decide when to buy from the grid and when to buy in the LEM.

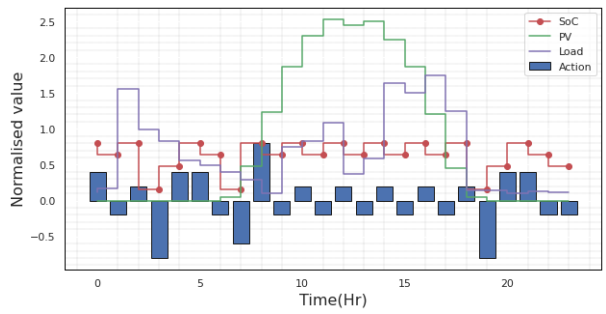

Fig. 4. Optimal trading strategy for a 24 hour period for House 7.

Fig. 4 shows the action performed over a 24 hour period considering the defined constraints for House 7. The model learned the best trading strategy to buy electricity from the grid or LEM when it is economically beneficial to perform that action. For instance, due to the high PV production between the hours of $10-18$ of Fig. 4, Actions 2 and 4 are mostly performed where PV is integrated with decision to trade in the LEM or the grid.

C. Economic benefit of trading in LEM

Fig. 5 shows the annual net benefit of each prosumer participating in energy trading either with the grid or in the LEM. This is analysed for when flexibility is provided with

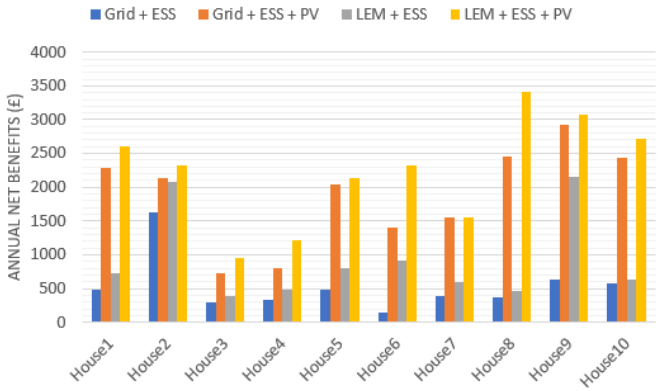

Fig. 5. The annual net benefit for ten houses with four action spaces; Grid+ESS, Grid+ESS+PV, LEM+ESS, LEM+ESS+PV.

ESS and PV or without for each prosumer/house. While the load profile and the PV capacities of each prosumer differ, the result illustrated that each house benefits is better-off while trading in the LEM than trading with the grid. For instance, for most houses, the average benefit of trading in the LEM over the grid with ESS is 35%. Furthermore, these net benefits increased when more sources of renewable energy are incorporated, reinforcing both the carbon and economic benefit of local energy consumption and trading. For instance, the average benefit increased to 54.6% when ESS and PV are incorporated during energy trading in the LEM.

V. Conclusion

This paper models a local energy trading strategy for 10 different houses equipped with different renewable energy resources including a combination of ESS and PV. In particular, a DQN-PER-based energy market algorithm is developed to optimise the utility derived by prosumers participating in a LEM as opposed to the grid. The benefits derived by the prosumers were analysed against the flexibility provided by the ESS and the incorporation of PV sources. The result showed the prosumers are better off trading energy in the LEM over the grid with an average benefit of 35%. These benefits increased when more sources of renewable energy are incorporated. The future work will focus on adapting the developed algorithm to the dynamic environment of prosumers with electric vehicles

to further optimise their utility of investing in renewable energy sources.

REFERENCES

[1] World Energy Council, "World energy trilemma index," 2017, available at https://www.worldenergy.org/assets/downloads/Energy-Trilemma-Index-2017-Report.pdf, Accessed: 2021-07-03.

[2] O. Jogunola, M. Hammoudeh, K. Anoh, and B. Adebisi, "Distributed ledger technologies for peer-to-peer energy trading," in *Electric Power and Energy Conference (EPEC)*. IEEE, 2020, pp. 1–6.

[3] O. Jogunola, B. Adebisi, A. Ikpehai, S. I. Popoola, G. Gui, H. Gačanin, and S. Ci, "Consensus algorithms and deep reinforcement learning in energy market: A review," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4211–4227, 2021.

[4] T. Baroche, P. Pinson, R. L. G. Latimier, and H. B. Ahmed, "Exogenous cost allocation in peer-to-peer electricity markets," *IEEE Transaction Power System*, vol. 34, no. 4, pp. 2553–2564, 2019.

[5] H. Almasalma, S. Claeys, and G. Deconinck, "Peer-to-peer-based integrated grid voltage support function for smart photovoltaic inverters," *Applied Energy*, vol. 239, pp. 1037–1048, 2019.

[6] A. Lüth, J. M. Zepter, P. C. del Granado, and R. Egging, "Local electricity market designs for peer-to-peer trading: The role of battery flexibility," *Applied energy*, vol. 229, pp. 1233–1243, 2018.

[7] O. Jogunola, B. Adebisi, K. Anoh, A. Ikpehai, M. Hammoudeh, and G. Harris, "Multi-commodity optimisation of peer-to-peer energy trading resources in smart grid," *Journal of Modern Power Systems and Clean Energy*, 2021.

[8] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, and G. Strbac, "Deep reinforcement learning for strategic bidding in electricity markets," *Transactions on Smart Grid*, vol. 11, no. 2, pp. 1343–1355, 2019.

[9] S. Wu, W. Hu, Z. Lu, Y. Gu, B. Tian, and H. Li, "Power system flow adjustment and sample generation based on deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1115–1127, 2020.

[10] Y. Zhou, B. Zhang, C. Xu, T. Lan, R. Diao, D. Shi, Z. Wang, and W.-J. Lee, "A data-driven method for fast ac optimal power flow solutions via deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1128–1139, 2020.

[11] T. Chen and W. Su, "Indirect customer-to-customer energy trading with reinforcement learning," *IEEE Transaction Smart Grid*, vol. 10, no. 4, pp. 4338–4348, 2018.

[12] T. Chen and W. Su, "Local energy trading behavior modeling with deep reinforcement learning," *IEEE Access*, vol. 6, pp. 62 806–62 814, 2018.

[13] T. Chen and S. Bu, "Realistic peer-to-peer energy trading model for microgrids using deep reinforcement learning," in *PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 2019, pp. 1–5.

[14] Y. Liu, D. Zhang, C. Deng, and X. Wang, "Deep reinforcement learning approach for autonomous agents in consumer-centric electricity market," in *International Conference on Big Data Analytics (ICBDA)*. IEEE, 2020, pp. 37–41.

[15] C. Hu, M. Kuklani, and P. Panek, "Accelerating reinforcement learning with prioritized experience replay for maze game," *SMU Data Science Review*, vol. 3, no. 1, p. 8, 2020.

[16] F. M. Uriarte, "Pecan street project field data," 2014, available at https://sites.google.com/site/fabianuriarte/downloads, Accessed: 2019-05-06.

[17] B. V. Mbuwir, M. Kaffash, and G. Deconinck, "Battery scheduling in a residential multi-carrier energy system using reinforcement learning," in *IEEE SmartGridComm*. IEEE, 2018, pp. 1–6.

[18] C. Hau, K. K. Radhakrishnan, J. Siu, and S. K. Panda, "Reinforcement learning based energy management algorithm for energy trading and contingency reserve application in a microgrid," in *PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. IEEE, 2020, pp. 1005–1009.

[19] Google, "Welcome to colaboratory," 2020, available at https://colab.research.google.com/, Accessed: 2020-11-16.