Decentralized Microgrid Energy Management: A Multi-agent Correlated Q-learning Approach

Hao Zhou, and Melike Erol-Kantarci, Senior Member, IEEE

School of Electrical Engineering and Computer Science

University of Ottawa

Emails:{hzhou098, melike.erolkantarci}@uottawa.ca

Abstract—**Microgrids (MG) are anticipated to be important players in the future smart grid. For proper operation of MGs an Energy Management System (EMS) is essential. The EMS of an MG could be rather complicated when renewable energy resources (RER), energy storage system (ESS) and demand side management (DSM) need to be orchestrated. Furthermore, these systems may belong to different entities and competition may exist between them. Nash equilibrium is most commonly used for coordination of such entities however the convergence and existence of Nash equilibrium can not always be guaranteed. To this end, we use the correlated equilibrium to coordinate agents, whose convergence can be guaranteed. In this paper, we build an energy trading model based on mid-market rate, and propose a correlated Q-learning (CEQ) algorithm to maximize the revenue of each agent. Our results show that CEQ is able to balance the revenue of agents without harming total benefit. In addition, compared with Q-learning without correlation, CEQ could save 19.3% cost for the DSM agent and 44.2% more benefits for the ESS agent.**

Index Terms—**Energy management, energy trading, correlated Q-learning, microgrid, smart grid.**

I. INTRODUCTION

Microgrids (MG) are becoming promising solutions to enhance the efficiency, resilience and flexibility of future smart grid [1]. Due to the integration of renewable energy resources and energy storage system (ESS), MGs are able to share their stored energy with each other to enhance reliability [2]. Resilient MGs are even expected to perform well under a catastrophic event to serve critical services as envisioned in [3]. The EMS of MG, which is the key of MG operation, could be centralized or decentralized. Centralized EMS needs to deal with the global information of MG, which will increase the complexity [4]. Therefore decentralized EMS is regarded as the future trend despite its control related challenges.

Meanwhile, the recent years have witnessed the great potential of AI techniques, which provides a new opportunity to improve the MG operation [5]–[7]. As a model-free algorithm, Q-learning can avoid the complexity of building a detailed optimization model. Q-learning is used in [5] to build a decentralized EMS in a MG where the benefits of agents is balanced by Nash equilibrium. [6] extended the reinforcement learning to MGs, which aimed to minimize the power loss using Bayesian reinforcement learning. In [7], the authors focus on the energy trading between MGs, where reinforcement learning is combined with Stackelberg game to find the Nash equilibrium.

In practice, the agents that control generation, storage, demand in a MG may belong to different entities, and they may be competing with each other for maximizing their revenues. In such a case, the coordination of agents are crucial for proper MG operation. In the literature, Nash equilibrium is widely used to coordinate the operation of agents [5], [7]–[9]. In [5], the interaction between agents became optimal when Nash equilibrium is reached, which means no agent can improve its expected reward by changing current strategy. In [9], Nash equilibrium is found in a distributed way, where no central controller is needed. However, Nash equilibrium is not guaranteed to converge or reach a global-optimal result, and some games have multiple Nash equilibriums. [9], [10]. In some cases, the Nash equilibrium is not unique or do not even exist [11].

To this end, we use the correlated equilibrium to coordinate agents, which is more general than Nash equilibrium. Firstly, the correlated equilibrium allows for dependencies among agents' strategies, while the actions in Nash equilibrium should be independent. Secondly, compared with iterative method which is generally used for finding Nash equilibrium, correlated equilibrium could be easily found by linear programming. The convergence and existence of correlated equilibrium is further proved in [10].

The main contribution of this paper is that we propose a multi-agent based correlated Q-learning (CEQ) algorithm for MG energy management which is implemented in a decentralized manner. The simulation results show that CEQ is capable of coordinating agents. The result shows the DSM agent could save as much as 19.3% of cost, and the ESS agent could earn 44.2% more benefit. Moreover, we compare the result with centralized EMS, where all agents belong to one aggregator, and we get the same total revenue which implies CEQ maintains the benefits of a centralized algorithm.

The rest of this paper is organized as follows. Section II presents the related work. Section III introduces our community MG system model, and Section IV introduces the proposed CEQ algorithm. Section V shows simulation results and finally, Section VI concludes the paper.

II. RELATED WORK

Energy management is the key for efficient operation of MGs. A controller coordinates the operation of different agents and optimizes the overall efficiency. With the integration of

RER, ESS, DSM and other agents, the complexity of MG energy management increases greatly. Based on AI techniques, learning-based methods become promising solutions for MG energy management. These have been investigated in a few recent studies which are summarized below [12]–[16].

In [12], fuzzy Q-learning is used for an isolated MG with multiple agents where the agents share their state variables for coordination. [13] uses a decision tree to chose actions for MG agents where the large training episodes are divided into small pieces, and Q-table is transferred between small episodes to speed convergence. Reinforcement learning and deep neural network is combined in [14] to conduct the energy management of multiple MGs. Furthermore, CEQ is used for dynamic transmission control of sensor networks in [15], where sensors learn the correlated equilibrium policy independently. In [16], CEQ is applied for smart generation control of power grids, where each area has one independent generation agent.

Considering the potential advantage of CEQ, this paper extends the application of CEQ for smart grid and MG integration. Different than [16], this paper mainly investigates the CEQ for MG energy management. In [16], the reward function is the same for each agent and the objective is solely generation control. However, the interactions between agents are more complicated in MGs since generation, storage and demand is involved. Therefore, one needs to define the reward function for each agent based on the their own characteristics. To the best of our knowledge, this is the first time CEQ is used for MGs.

III. COMMUNITY MICROGRID SYSTEM MODEL

In this section, we will introduce the demand side management model, the PV model and the energy storage system model. Based on mid-market rate, an improved energy trading model is built.

A. Demand Side Management Model

The user demand is divided as crucial load and deferrable load. The demand side management (DSM) agent does not interfere with the crucial load, e.g., lighting and cooking devices. However deferrable loads, such as a dish washer or a water heater, can be deferred to another time by the DSM agent.

For n sets of deferrable devices, the state is described as:

$$\vec{a}_t = [a_{t,1}, a_{t,2}, \ldots, a_{t,i}, \ldots, a_{t,n}] \tag{1}$$

where $a_{t,i} = 1$ represents turning on the devices in set i at time t; and $a_{t,i} = 0$ represents turning off.

The power demand of DSM agent at time t is:

$$P_t^{DSM} = \sum_{i=1}^{n} P_i a_{t,i} \tag{2}$$

where P_i is the power of deferrable devices set i. We assume an average power consumption for the duration $[t, t+1]$.

The deferrable devices must be serviced before a certain time limit, which means devices can not be deferred longer than this limit. The waiting time of devices is described as:

$$\vec{w}_t = [w_{t,1}, w_{t,2}, \ldots, w_{t,i}, \ldots, w_{t,n}] \tag{3}$$

where $w_{t,i} = 1$ represents the devices set i that is still under service at time t; $w_{t,i} = 0$ represents this devices set has been serviced or its turn has not come. We assume that if one device reaches its time limit and it has not been serviced, then it will be turned on mandatorily.

B. PV Model

In this paper, we assume the PV power can be predicted with an acceptable error [17].

$$P_t^{PV} = \hat{P}_t^{PV} \tag{4}$$

where $\hat{P}_t^{PV}$ is the predicted PV power.

C. Energy Storage System Model

In this research, we consider centralized ESS in the community MG. We use q_t to denote the state of ESS at time t. There are two discharging levels, denoted as q_t equals 0.5 and 1, to enhance flexibility. This means the ESS can be either fully discharged or half discharged. The power of ESS is:

$$P_t^{ESS} = P^{char} q_t \tag{5}$$

$$q_t = \begin{cases} -1 & charge \\ 0 & unchanged \\ 0.5 & discharge \\ 1 & discharge \end{cases} \tag{6}$$

where P^{char} is a constant charging power.

The state of charge (SOC) of ESS is updated according to:

$$SOC_{t+1} = SOC_t - \frac{P^{char}}{C^{ESS}} q_t \tag{7}$$

where C^{ESS} is the capacity of ESS.

D. Energy Trading Model

Our energy trading model is presented in Fig.1. According to our model, DSM agent could choose ESS or main grid as its energy source. ESS agent could use PV or grid power to charge, and it can sell its energy to the DSM agent or the main grid. We make the following assumptions to build the energy trading model:

Assumption 1: Compared with the main grid, PV power has a lower price: $p^{PV} < p_t^{bgrid}$. Considering the rationality of main grid, $p_t^{sgrid} < p_t^{bgrid}$.

Assumption 2: Surplus PV power is only available when PV power is more than the crucial load. Note that energy trading of crucial load is out of the scope of this paper.

Assumption 3: We assume ESS is unable to charge and discharge at the same time.

Assumption 4: If DSM is unable to consume all the ESS energy, ESS could sell the rest energy to main grid.

Fig. 1. Energy trading model.

TABLE I
STRATEGY COMBINATION OF [PLAYERS

Value of middle price		ESS agent	
		Cooperation	Threat
DSM Agent	Cooperation	$0.5p^{sgrid} + 0.5p^{bgrid}$	$0.25p^{sgrid} + 0.75p^{bgrid}$
	Threat	$0.75p^{sgrid} + 0.25p^{bgrid}$	Cooperation breaks

The mid-market rate has been generally used in P2P energy trading research as in [18]. We use p_t^{mid} to represent the price of ESS selling electricity to DSM.

$$p_t^{mid} = \frac{p_t^{sgrid} + p_t^{bgrid}}{2} \tag{8}$$

where p_t^{sgrid} and p_t^{bgrid} represent the price of selling power to the main grid and buying power from the main grid separately.

However, in a competitive situation, it is reasonable to assume one agent may take risks for a higher profit, which is denoted as a threat strategy. The strategy combination and results are presented in Table 1. If one agent chooses a threat strategy and other agent still cooperates, the former agent will earn more benefit. If both of them choose threat strategy, then cooperation breaks up, both agents will exchange power with main grid. In this paper, we will use correlated equilibrium to coordinate the actions of agents, avoid conflict and balance the revenue.

E. Problem Formulation

Both ESS and DSM agent make decisions independently to optimize their own operation. The optimization objective of the ESS agent is maximizing its revenue:

$$max(P^{char} \sum_{t=1}^{T} q_t p_t^{clear}) \tag{9}$$

where T is simulation period; p_t^{clear} is the price of ESS agent buying/ selling electricity, which is known as the market clearing price [13]. Clearing price denotes an ideal situation, where the demand equals the supply and no shortage or surplus exist in the market. P^{char} and q_t have been defined in equation (5).

On the other hand, the optimization objective of DSM agent is minimizing cost:

$$min(\sum_{t=1}^{T} P_t^{DSM} p_t^{buy}) \tag{10}$$

where p_t^{buy} is the price of buying power at time t.

The optimization has to obey following constraints:

$$P_t^{PV} + P_t^{grid} + P_t^{ESS} = P_t^{DSM} \tag{11}$$

$$SOC_{min} \leq SOC_t \leq SOC_{max} \tag{12}$$

$$a_{t,i} \leq w_{t,i} \tag{13}$$

Equation (11) denotes the energy balance constraint, where P_t^{grid} is the power from the main grid. Equation (12) is the SOC constraint, where SOC_{min} and SOC_{max} are lower and upper bound of ESS. Equation (13) is the DSM constraint, which means only devices that have not been serviced could be turned on.

IV. CORRELATED Q-LEARNING ALGORITHM

In this paper, we propose a correlated Q-learning (CEQ) based algorithm to coordinate the operation of agents. CEQ is a multi-agent reinforcement learning algorithm. The coordination between agents is achieved by exchanging the state-action value matrix, which means it could be implemented in a decentralized way. We will introduce the system state, actions, rewards and correlated equilibrium in this section.

A. State and Actions

The system state is defined as:

$$s = \{t, SOC, \vec{w}\} \tag{14}$$

The actions of DSM and ESS agent are:

$$a^{DSM} = \{y, \vec{a}\} \tag{15}$$

$$a^{ESS} = \{x, u\} \tag{16}$$

where y belongs to a set of two choices: buying electricity from ESS or main grid; x belongs to a set of two choices: selling power to main grid or DSM, u is the set of q_t in eq. (5).

The Q value spaces of DSM and ESS agent are:

$$Q_{DSM} = \{a^{ESS}, a^{DSM}, t, SOC, \vec{w}\} \tag{17}$$

$$Q_{ESS} = \{a^{ESS}, a^{DSM}, t, SOC, \vec{w}\} \tag{18}$$

B. Reward

The cost of ESS agent includes buying power from PV or main grid, and the benefit comprises of selling power to grid or DSM. Meanwhile, DSM agent aims to minimize the cost by adjusting operation time of deferrable devices and choose a lower price energy source. The reward can be formulated for the following two cases that correspond to ESS charging or discharging.

1) At time t, if ESS agent chooses to charge, the instant reward is:

$$r_t^{ESS} = P_t^{ESS}(\alpha p_t^{PV} + (1-\alpha)p_t^{bgrid}) \qquad (19)$$

where α denotes the proportion of buying power from PV $(0 \leq \alpha \leq 1)$.

ESS agent will prefer PV power because of a lower price, but surplus PV power may not be enough for ESS charge. As a liner program problem, we give the following equation of α:

$$\alpha = \frac{P_t^{PV} - P_t^{crl}}{\left|P_t^{ESS}\right|} \qquad (20)$$

where P_t^{crl} is the crucial load.

Meanwhile, the DSM can only get power from the main grid. The reward of DSM agent at time t is:

$$r_t^{DSM} = -P_t^{DSM}p_t^{bgrid} \qquad (21)$$

2) If ESS agent chooses to discharge, the reward of ESS agent is:

$$r_t^{ESS} = P_t^{ESS}(\beta p_t^{mid} + (1-\beta)p_t^{sgrid}) \qquad (22)$$

where β denotes the proportion of selling power to DSM $(0 \leq \beta \leq 1)$. According to equation (8), the ESS will try to sell power to DSM agent for a higher price. Note that DSM may be unable to buy all ESS power, therefore we give the following equation:

$$\beta = \frac{P_t^{DSM}}{P_t^{ESS}} \qquad (23)$$

The reward of DSM agent is:

$$r_t^{DSM} = -P_t^{DSM}(\gamma p_t^{mid} + (1-\gamma)p_t^{bgrid}) \qquad (24)$$

Similarly, denoting the proportion of buying power from ESS by:

$$\gamma = min(\frac{P_t^{ESS}}{P_t^{DSM}}, 1) \qquad (25)$$

C. Correlated Equilibrium

The aim of ESS and DSM agent are both to maximize their reward in a simulation epoch. For the agent i, with the initial state s_0, the expected accumulated discounted reward under a policy π is:

$$V_i^\pi(s) = E_\pi[\sum_{n=0}^{\infty} \theta^n r_i(s_n, a_n)|s = s_0] \qquad (26)$$

where θ is the reward discount factor.

For a specific state s and action a, we define the state-action value:

$$Q_i^\pi(s,a) = r_i(s,a) + \theta \sum_{s' \in S} P(s'|s,a)V_i^\pi(s) \qquad (27)$$

Then we have the following relationship:

$$V_i^\pi(s) = \sum_{a \in A} \pi_x(a)Q_i(s,a) \qquad (28)$$

In the CEQ algorithm, agents chose a joint optimal action by exchanging Q value matrix, and we assume both agents know the possible actions of each other. The joint action of agents are chosen according to correlated equilibrium:

$$\sum_{a_{-i} \in A_{-i}} \pi_x(a_{-i}, a_i)Q_i(s, a_{-i}, a_i) \geq$$
$$\sum_{a_{-i} \in A_{-i}} \pi_x(a_{-i}, a_i)Q_i(s, a_{-i}, a_i') \qquad (29)$$

where the a_i and a_i' denote the action of agent i in correlated equilibrium and non-equilibrium, a_{-i} denotes the actions of agents except agent i, A_{-i} denotes the action set of agents except i. Meanwhile, equation (29) denotes that the joint-optimal action is chosen by exchanging Q-values, which means the private information of each agent could be well protected.

The policy is improved according to ϵ-greedy policy:

$$\pi_i(s) = \begin{cases} random & rand \leq \epsilon \\ equation(29) & rand > \epsilon \end{cases} \qquad (30)$$

Where ϵ is a small value between 0 and 1; $rand$ is a random number between 0 and 1.

It is worth noting that ESS will only choose eligible actions, e.g., the charge action is eliminated if SOC=1. The CEQ algorithm is summarized in Algorithm 1.

Algorithm 1 Correlated Q-learning

1: **Initialize:** microgrid and Q-learning parameters
2: **for** $j = 1$ to $episode$ **do**
3:   Reset state s
4:   **for** $t = 1$ to T **do**
5:     Calculate eligible a_t^{DSM}, a_t^{ESS} according to state s
6:     **if** $rand < \epsilon$ **then**
7:       Randomly choose a_t^{DSM}, a_t^{ESS}
8:     **else**
9:       Agents exchange current state-action matrix
10:       Find equilibrium by equation (29) and get optimal joint action a_t^{DSM}, a_t^{ESS}
11:     **end if**
12:     $[r_t^{DSM}, r_t^{ESS}]$ = Reward $(s, a_t^{DSM}, a_t^{ESS})$
13:     Update Q, t, SOC and $\vec{w}$
14:   **end for**
15: **end for**
16: **Output:** Optimal action sequence from $t = 1$ to T

V. SIMULATION RESULT

A. Parameter Settings

We use MATLAB for our simulations. We model the PV power and the crucial load as shown in Fig.2, which are extracted from [17]. Surplus PV power is available when PV power is higher than crucial load. Fig.3 shows the Time-of-Use (TOU) electricity price, which is generated according to winter TOU price of Ontario, Canada.

Fig. 2. PV power and crucial load.

Fig. 3. TOU electricity price.

TABLE II
PARAMETERS SETTINGS

Device Number	Operation time limit (Hours)	Average duration time (Hours)
1	[1, 8]	1
2	[7, 13]	1
3	[10, 17]	2
4	[15, 21]	1
5	[20, 5^{+24h}]	3

TABLE III
PARAMETERS SETTINGS

Parameters	value
Capacity of ESS: C^{ESS}	$120kWh$
ESS charge power: P_{char}	$30kW$
Price of PV selling electricity to ESS: p^{PV}	0.03 $/kWh$
Price of selling electricity to grid: p^{sgrid}	0.08 $/kWh$

Table 2 shows the operation time limits and the operation duration of five sets of deferrable devices. For each device, the power (kW) is an integer number generated randomly from the set [8, 14] which represents the average power consumed throughout the operation. We repeat the simulations for 30 runs and present the averaged values with 95% confidence intervals in plots. Table 3 shows the parameter settings of our simulations.

B. Energy Management with Correlated Q-learning

First, we set the initial SOC of ESS to 0.25, and correlated Q-learning is used to optimize the cost of ESS agent and profit of DSM agent. The result of average DSM cost and ESS profit are shown in Fig.4, where the cost decreases and profit increases through the iterations of the algorithm. It is observed that both agents converge.

Fig. 4. Average DSM cost and ESS benefit with iterations.

Fig. 5. Optimal microgrid energy balance under correlated Q-learning

Fig.5 presents how the optimal MG energy management works. Positive bars represent load, charging or electricity sold to grid, while negative bars represent discharging or surplus PV power. The negative bars and positive bars have the same total height at each hour, which means energy balance is maintained. For example, from 1:00 to 2:00, ESS agent discharges with a power 15kW, and DSM agent uses 10kW for operation, while the rest 5kW is sold to main grid. Due to

a lower charging price, ESS agent uses surplus PV power to charge from 11:00 to 15:00. Meanwhile, DSM agent uses ESS power for operation, and both agents benefit from cooperation.

Fig.6 shows how devices are deferred. Device 1 and 2 start immediately because they have already worked under the minimum TOU price in their time limit. Device 3 and 4 are deferred 5 and 2 hours later respectively for a lower TOU price. Although the TOU price is lower from 13:00 to 18:00, note that device 3 starts operation at 15:00, because the ESS uses surplus PV power for charging from 11:00 to 15:00, and the ESS is unable to discharge in this period. Device 5 are deferred to 1:00 of next day.

Fig. 6.   Deferred operation time of each device

C. Comparison under Different PV Power

In the following two sections, we will make a comparison of following three cases:

Case I: CEQ is used to coordinate two agents.

Case II: Q-learning without correlation (QLWC) is implemented, where each agent maximizes their own reward and ignore the other agent.

Case III: Q-learning with aggregator (QLWA) is conducted in this case, where all agents belong to one aggregator and maximize the total revenue. Considering the centralized optimization can usually get the global optimal result, we compare CEQ with QLWA to show that the distributed nature of CEQ does not harm total revenue.

The initial SOC of ESS is set to 0.25, and two cases is compared under different peak PV power. As shown in Fig.7 and Fig.8, the CEQ outperforms QLWC with a lower DSM cost and a higher ESS profit, respectively. The 100% Peak PV power is defined as Fig.2. At 100% PV power, CEQ saved 16.5% cost for DSM agent, and earn 25.5% more benefit for ESS agent. At 80% PV power, the proportion for DSM and ESS become 16.3% and 30.4%, respectively. CEQ earned at most 44.2% more benefit than QLWC when peak PV power is only 20%.

The main reason is the CEQ is able to coordinate the operation of DSM and ESS. ESS use PV power to charge, and sell the energy to DSM, where both of them benefit from cooperation. On the contrary, in QLWC, both agents try to maximize their own reward and cooperation breaks. Instead of buying energy from ESS agent, DSM agent chooses to buy electricity from main grid. Meanwhile, ESS agent uses PV power to charge, and sell the electricity to main grid.

Fig. 7.   Comparison of DSM cost under different PV power

Fig. 8.   Comparison of ESS profit under different PV power

D. Comparison under Different Initial SOC of ESS

In this section, we set the PV power to 100%, and compare three cases under different initial SOC of ESS. Note that the cost of intial SOC is not considered.

Fig. 9.   Comparison of DSM cost under different initial SOC

As shown in Fig. 9 and Fig. 10, the CEQ has a lower DSM cost and a higher ESS profit than QLWC. DSM agent uses ESS energy for operation, both agents benefit. At 0.75 initial SOC, the DSM agent in CEQ could save 19.3% cost, and ESS agent in CEQ could earn 16.3% more benefit than QLWC. After the DSM agent gets enough energy from ESS, the ESS will sell surplus energy to main grid and earn a profit. As a result, the DSM cost remain unchanged, but ESS profit could still increase.

Fig. 10. Comparison of ESS profit under different initial SOC

Furthermore, at 100% PV power, we compare the performance of CEQ with QLWA in Table 4. If we use the profit of CEQ minus the cost, then we will get the same total revenue with QLWA. The result proves CEQ did not harm the total benefit of whole MG.

TABLE IV
COMPARISON OF CEQ AND QLWC UNDER DIFFERENT INITIAL SOC ($)

Initial SOC		0.25	0.5	0.75	1.00
CEQ	Profit	10.55	12.95	15.35	17.75
	Cost	9.20	9.23	9.20	9.21
QLWA-Revenue		1.35	3.72	6.15	8.54

VI. CONCLUSION

In this paper, we propose a multi-agent correlated Q-learning algorithm for microgrid energy management. Compared with the Q-learning without correlation, we find that the proposed CEQ scheme could save at most 19.3% cost for DSM agent, and earn at most 44.2% more benefit for ESS agent. The simulation results show that CEQ is capable of successfully coordinating the operation of agents. In addition, compared with an aggregator case, CEQ is shown to maintain the same total benefit for the agents.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Yoldas, A. Onen, S. Muyeen, A. Vasilakosc, and I. Alan, "Enhancing smart grid with microgrids: Challenges and opportunities," Renewable and Sustainable Energy Reviews, vol. 72, pp. 205-214, May 2017.

[2] M. Erol-Kantarci, B. Kantarci, and H. T. Mouftah, "Reliable overlay topology design for the smart microgrid network," IEEE Network, vol. 25, pp. 38–43, Sep 2011.

[3] L. Wu, J. Li, M. Erol-Kantarci, and B. Kantarci, "An integrated reconfigurable control and self-organizing communication framework for community resilience microgrids," The Electricity Journal, vol. 30, pp.27-34, May 2017.

[4] L. Meng et al., "Microgrid supervisory controllers and energy management systems: A literature review," Renewable and Sustainable Energy Reviews, vol. 60, pp. 1263-1273, Jul 2016.

[5] E. Foruzan, L. Soh, and S. Asgarpoor, "Reinforcement Learning Approach for Optimal Distributed Energy Management in a Microgrid," IEEE Trans Power Syst., vol. 33, pp. 5749-5758, Sep 2018.

[6] M. Sadeghi, and M. Erol-Kantarci, "Power Loss Minimization in Microgrids Using Bayesian Reinforcement Learning with Coalition Formation", in Proc. of IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-6, Istanbul, Turkey, Sep 2019.

[7] H. Wang, T. Huang, X. Liao, H. Abu-Rub, and G. Chen, "Reinforcement Learning in Energy Trading Game Among Smart Microgrids," IEEE Trans. Ind. Electron., vol. 63, pp. 5109-5118, Aug 2016.

[8] K. Luo and W. Shi, "Distributed Coordination Method of Microgrid Economic Operation Optimization Based on Multi-Agent System" ,in Proc. of Int. cof. on Power System Tech., pp. 1-6, Chengdu, China, Oct 2014, .

[9] J. Zeng, Q. Wang, J. Liu, J. Chen, and H. Chen, "A Potential Game Approach to Distributed Operational Optimization for Microgrid Energy Management With Renewable Energy and Demand Response," IEEE Trans. Ind. Electron., vol. 66, pp. 4479-4489, Jun 2019.

[10] A. Greenwald, K. Hall, and M. Zinkevich, "Correlated Q-learning," Dept. of Comput. Sci., Brown University, USA, Tech. Rep.,CS-05-08, 2005.

[11] T. Eisele, "Nonexistence and Nonuniqueness of Open-Loop Equilibria in Linear-Quadratic Differential Games," Journal of optimization theory and applications, vol. 37, pp. 443–468, Aug 1982

[12] P. Kofinas, A.I. Dounisb, and G.A. Vourosa, "Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids," Applied Energy, vol. 219, pp. 53-67, Jun 2018.

[13] T. Levent, P. Preux, E. L. Pennec, J. Badosa, G. Henri, and Y. Bonnassieux, "Energy Management for Microgrids: a Reinforcement Learning Approach," in Proc. of IEEE PES Innovative Smart Grid Technologies Europe, Bucharest, Romania, Sep 2019.

[14] Y. Du, and F. Li, "Intelligent Multi-microgrid Energy Management based on Deep Neural Network and Model-free Reinforcement Learning," IEEE Trans. Smart grid, vol. 11, Mar 2020.

[15] J. W. Huang, Q. Zhu, V. Krishnamurthy, and T. Basar, "Distributed Correlated Q-Learning for Dynamic Transmission Control of Sensor Networks," in Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Dallas, TX, USA, Mar 2010.

[16] T. Yu, H. Z. Wang, B. Zhou, K. W. Chan, and J. Tang, "Multi-Agent Correlated Equilibrium Q(λ) Learning for Coordinated Smart Generation Control of Interconnected Power Grids," IEEE Trans. Power Syst., vol. 30, pp. 1669–1679, Jul 2015.

[17] European Network of Transmission System Operators for Electricity (ENTSO-E), [Online]. Available: https:// transparency. entsoe.eu/dashboard/show

[18] C. Long, J. Wu, C. Zhang, L. Tomas, M. Cheng, N. Jenkins, "Peer-to-peer energy trading in a community microgrid," in Proc. of IEEE Power Energy Gen. Meeting (PES), pp. 1-5, Chicago, IL, USA., Jul 2017.