Artificial Intelligence Based Smart Energy Community Management: A Reinforcement Learning Approach

Suyang Zhou, Zijian Hu, Wei Gu, *Senior Member*, *IEEE*, Meng Jiang, and Xiao-Ping Zhang, *Senior Member, IEEE*

***Abstract*—This paper presents a smart energy community management approach which is capable of implementing P2P trading and managing household energy storage systems. A smart residential community concept is proposed consisting of domestic users and a local energy pool, in which users are free to trade with the local energy pool and enjoy cheap renewable energy while avoiding the installation of new energy generation equipment. The local energy pool could harvest surplus energy from users and renewable resources, at the same time it sells energy at a higher price than Feed-in-Tariff (FIT) but lower than the retail price. In order to encourage the participation in local energy trading, the electricity price of the energy pool is determined by a real-time demand/supply ratio. Under this pricing mechanism, retail price, users and renewable energy could all affect the electricity price which leads to higher consumers' profits and more optimized utilization of renewable energy. The proposed energy trading process was modeled as a Markov Decision Process (MDP) and a reinforcement learning algorithm was adopted to find the optimal decision in the MDP because of its excellent performance in on-going and model-free tasks. In addition, the fuzzy inference system makes it possible to use Q-learning in continuous state-space problems (Fuzzy Q-learning) considering the infinite possibilities in the energy trading process. To evaluate the performance of the proposed demand side management system, a numerical analysis is conducted in a community comparing the electricity costs before and after using the proposed energy management system.**

***Index Terms*—Artificial intelligence, distributed management, fuzzy Q-learning, microgrid, reinforcement learning.**

I. Introduction

THE smart grid aims at adopting advanced information, digital and intelligent technologies to facilitate the accommodation of renewable energy and Distributed Energy Resources (DERs) and improve the reliability and economic efficiency of the electric power network. In recent years, significant peak-time demand enlarges the imbalance between electricity generation and load [1], [2] because of the increasing penetration of Distributed Generators (DGs), electrification of heating and transportation (i.e. Electric Vehicles (EVs) and Heat Pumps (HPs)). The management of DGs and DERs are becoming increasingly important for grid operators. Meanwhile, DG owners believe it difficult to recover their investment in a reasonable period, because their governments turned down the Feed-in-Tariff (FIT). The decrease of the FIT will prolong the pay-back period of DGs and deteriorate the motivations for further investment on DGs, which will correspondingly have an inactive impact on achieving the greenhouse gas reduction target set by the EU and other countries. Therefore, it is necessary and urgent to explore new business models or energy management approaches for DGs and other DERs to maintain the interests for installing DGs.

To help DG and DER owners improve their economic benefits, a number of proposed researches focused on Demand Side Management (DSM) [3]–[6]. Previous studies carried out on DSM have demonstrated the potential benefits of DSM on augmenting the usage of DGs and decreasing the bills of electricity end users [7].

Optimal load and DER scheduling is considered as one of the most efficient applications for DSM. An autonomous and distributed demand-side energy management system was presented in [8]. In the proposed system, a household energy consumption scheduling game was formulated as a game in game theory, where the optimal economic performance is expected to be achieved at the Nash equilibrium. In addition, the distributed strategy only needs the users' current load and tariffs which are beneficial as the communication requirement and then the pressure can be relaxed. Authors in [9] described a real-time optimization approach with the objective of minimizing daily energy-consumption and they incorporated this model into a simple linear optimization problem. [10] and [11] also proposed the game-theoretic energy management approach but having it implemented by algorithms different from [2].

An alternative approach for DSM is direct load control (DLC). In [12], dynamic programing (DP) was used to obtain the optimal DLC scheme with operating cost being the lowest. A distributed direct load control scheme for residential demand response (DR) built on a two-layer communication-based control architecture, is proposed in [13], [14] investigated the

Manuscript received August 24, 2018; revised November 15, 2018; accepted Janurary 3, 2019. Date of publication March 30, 2019; date of current version February 19, 2019. This work was supported by the National Natural Science Foundation of China (No. 51807024).

S. Y. Zhou, Z. J. Hu and W. Gu (corresponding author, email: wgu@seu.edu.cn) are with the School of Electrical Engineering, Southeast University, Nanjing, China.

M. Jiang is with the Department of Computer Science and Engineering, University of Notre Dame, USA.

X. P. Zhang is with the School of Electronic, Electrical and System Engineering, University of Birmingham, U.K.

DOI: 10.17775/CSEEJPES.2018.00840

thermal comfort level during a DLC event and introduced a detailed multi-level linear modeling of thermal sensation. However, the concept of a DLC scheme is not widely accepted by the end users, especially the residential users. [15] reported that only 13% customers were willing to participate in a DLC program based on the survey of 1,499 households in one state of Australia. The main reasons for reluctance in participating in a DLC program are the distrust in energy companies and the potential risk of individual privacy disclosure. Thus, it is essential to take cyber security issues into account when designing a DLC scheme, and it may also be advisable to avoid the centralized control of users' privacy sensitive data, such as clothes drying and household heating.

Meanwhile, the Time-of-Use (TOU) tariffs have been introduced by a number of energy companies in the US, EU and Asia [16], [17]. Many researches on DSM, incorporated with TOU tariffs, have been carried out. [18] showed that combining DSM and TOU tariffs could significantly improve the system operation security and reduce the costs and emissions of energy systems at high renewable penetration levels. [19] developed a decision-support system that helps users save energy by recommending optimal run-time schedules for home appliances based on constraints and TOU tariffs. The Demand Response (DR) was modeled considering both TOU and Emergency Demand Response Program (EDRP) methods in [20], and the optimal incentives for combined TOU and EDRP programs are determined. From the aforementioned studies, it is obvious that the TOU tariff does bring benefits to both electricity end users and network operators, which means pricing mechanism will directly benefit the energy consumption behavior and network operation. Recently, researchers proposed certain new trading and pricing concepts, such as Peer-to-Peer (P2P) trading and a local energy trading pool, for electricity end users.

The P2P energy trading mechanism enables the energy trading of surplus energy among the users. Along with the incremental penetration of DGs and DERs, the role of electricity end users are transferring from pure consumer to prosumer. The P2P trading allows prosumers to sell their surplus energy and gives other consumers the option of buying the energy from prosumers, which can increase the energy market liquidity and ideally reduce the energy price. [21] proposed a P2P energy trading system between two sets of electric vehicles. The results showed that the proposed system can significantly reduce the impact of the charging process on the power system during peak time. [22] developed a Multi-Agent System (MAS) based on a day-ahead control algorithm for communities with P2P trading capability, where homes could react with environmental changes and trade with other agents. In [23], Professor Wu and his group proposed three market paradigms (bill sharing, mid-market rate and auctions based on pricing strategy) and applied them to a community microgrid as an example. The feasibility of P2P energy trading to reduce consumer energy costs and to increase income for DER producers was accessed in [23]. For P2P trading, the market participants would need to make the decision of how much energy they want to buy/sell and during what time periods. The decision-making process is complex and requires high computation power; this would be an obstacle to promoting P2P trading for domestic users. Thus, it is important to explore the efficient approaches of controlling the DERs for the domestic users to facilitate the P2P trading mechanism.

This paper proposes a P2P trading system for a domestic community with a local energy pool. The proposed P2P trading system allows domestic users to trade their surplus energy into the energy pool to obtain more benefits than the FIT tariff. In our proposed system, pure consumers can also buy energy from the pool with a relatively lower price than the retail market price. Users owning an Energy Storage System (ESS) can obtain extra benefits via arbitrage trading. Some researchers have done studies about P2P energy trading in communities, such as the P2P trading mechanism proposed in [23]. Compared to the approach they proposed, we hope to improve the sharing of renewable energy and reduce consumers' energy costs, which is different to the motivations in [23]. In addition, P2P trading mechanisms in [23] are based on a detailed model of business and energy exchange pricing. In our paper, we focus on the model-free algorithm for speeding up the optimization. We model the pricing mechanism as a game between the demand and surplus ratios in a simplified way, and we propose a novel fuzzy Q learning algorithm to find optimal solutions that support the distributed operation of market participants in the P2P trading system. The main contributions of this paper are as follows:

This paper presents a P2P trading infrastructure for the residential community that aims to maximize the sharing of surplus energy and facilitate the operational efficiency of ESS.

A reinforcement learning (Fuzzy Q-learning) algorithm is proposed to improve the decision-making process for the market participants in a model-free way. Detailed models for market participants are no longer needed because the Q-leaning algorithm can solve the optimal decision problems without requiring a model, and consequently the computation time of the decision making process will be significantly reduced.

The rest of this paper is organized as follows: Section II describes the smart community concept that is capable of conducting P2P trading and presents the model of price and storage. Section III proposes a Fuzzy Q-learning Algorithm and the MDP framework for the community. Section IV illustrates the effectiveness of the proposed algorithm with three case studies, and the conclusion is drawn in Section V.

II. PROBLEM STATEMENT

A. Distributed Energy Sharing Community Description

This paper depicts a residential community framework consisting of one local energy pool, a set of smart homes and a set of non-intelligent homes. The smart homes in the community are equipped with ESS, which can help the smart homes sell their surplus energy to the energy pool and buy the energy back from the pool with consideration of the real-time energy price. For the non-intelligent homes, they can buy the energy from the energy pool with relatively lower prices than the retail market price, and can also buy the energy from the retail market when there is not sufficient energy available

in the energy pool. If the non-intelligent homes are installed with DGs such as a solar system, they can sell their surplus energy to the energy pool with higher prices than the FIT tariff. The action of buying/selling energy from/to the energy pool are determined by the decentralized agents. The schematic diagram of the energy sharing community structure is shown in Fig. 1. Such a design allows smart homes the freedom to sell electricity to obtain extra benefits rather than merely obtaining the compensation through the FIT. The non-intelligent users can also obtain benefits from the relatively lower price than the retail market by participating in the P2P scheme.

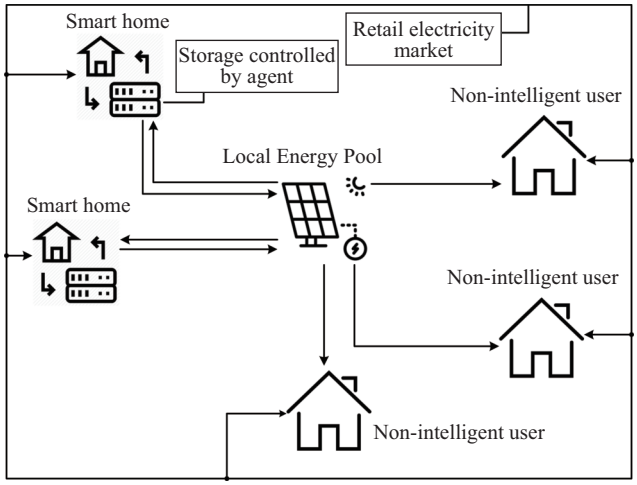


Fig. 1. Structure of energy sharing community.

B. Local Energy Pool

The local energy pool is designed for collecting surplus energy from users and DGs, and providing energy at a lower price than the retail market price or Real Time Price (RTP). The energy within the pool comes from two sources: one is the energy sold by smart homes and the other is from the surplus energy from DGs. All market participants of the energy pool are under the volunteer principle and will accept the market price decided by the surplus-to-demand ratio. [13] presented an electricity pricing model considering the real-time retail market price and aggregated local energy profile. However, in such a model, the price could be higher than the retail market price when the aggregated profile is positive (the supply is higher than the demand). This will obviously reduce the motivation of smart homes in participating in P2P trading. To facilitate the participation in the P2P trading system, the surplus-vs-demand ratio is taken into account. The pricing model is presented as follows:

$$p(t,\gamma(t)) = r(t) - q(t,\gamma(t)) \tag{1}$$

s.t.

$$q(t,\gamma(t)) = a(t)\cdot\gamma(t)^2 + b(t)\cdot\gamma(t) \tag{2}$$

$$q(t,\gamma(t)) + p^F < r(t) \tag{3}$$

$$\gamma(t) = \frac{e_m(t) + s_{\mathrm{pv}}(t) + s_m(t)}{s_{\mathrm{n}}(t)}, \tag{4}$$

where p denotes the real-time market price in the community and $\gamma(t)$ denotes the ratio of the energy in the pool and demand at time t. $a(t)$ and $b(t)$ are time-dependent non-negative parameters. $s_m(t)$ denotes the amount of energy that smart users sell to the pool at time t, $e_m(t)$ is the rest of the energy in the pool before time t, $s_{\mathrm{pv}}(t)$ is the solar energy at time t and $s_{\mathrm{n}}(t)$ denotes the aggregated demand amount of users that are requesting energy from the pool. p^{F} denotes the Feed-in-Tariff (FIT). Equation (2) has been discussed in [16]. Electricity prices given by this model are lower or at least equal to the real-time price but higher than the FIT.

C. Smart Home Agent

The smart home agent is responsible for managing the charging/discharging operation of the ESS considering the energy pool price and the neighborhood demand.

In the proposed P2P trading system, the ESS is the key controllable equipment that enables users to get involved in the P2P trading system. The model and constraints of the ESS are as follows [15], [17], [18]:

$$E^{\mathrm{Bt}}_{(t+\Delta t)} = \begin{cases} E^{\mathrm{Bt}}_t + P^{\mathrm{b}}_t\cdot\eta^{\mathrm{c}}\cdot\Delta t & if \quad P^{\mathrm{b}}_t \geq 0 \\ E^{\mathrm{Bt}}_t + \frac{P^{\mathrm{b}}_t}{\eta^{\mathrm{d}}}\cdot\Delta t & if \quad P^{\mathrm{b}}_t < 0 \end{cases} \tag{5}$$

s.t.

$$P^{\mathrm{b}}_{\mathrm{d}} \leq P^{\mathrm{b}}_t \leq 0 \quad \text{if} \quad P^{\mathrm{b}}_t < 0 \tag{6}$$

$$0 \leq P^{\mathrm{d}}_t \leq P^{\mathrm{b}}_c \quad \text{if} \quad P^{\mathrm{b}}_t \geq 0 \tag{7}$$

$$0 \leq E^{\mathrm{Bt}}_t \leq E^{\mathrm{Bt}} \tag{8}$$

where E^{Bt}_t is the capacity of Battery Energy Storage system (BESS) at time t, P^{b}_t is the charging/discharging rate of BESS at time t, η^{c} is the charging efficiency, η^{d} is the discharging efficiency, P^{b}_c is the BESS maximum charging rate, $P^{\mathrm{b}}_{\mathrm{d}}$ is the BESS maximum discharging rate and E^{Bt} is the maximum BESS capacity.

With regard to the household ESS, the corresponding agent is required to formulate the power plan in real time, which aims to achieve greater profits for the users. There are primarily three factors that will affect the decision-making for the ESS: battery capacity, public retail price and the neighborhood price in the community.

Discharge: If the remaining energy could not meet user's requirements, the agent would buy energy from the retail market or local energy pool. Otherwise, the ESS may discharge more electricity than demand to participate in energy trading. Energy traded to the energy pool would reduce the ratio of supply to demand in equation (4) and lead to electricity depreciation in the local energy pool.

Charge: Under this circumstance, the agent has to buy electricity from the retail market or local energy pool. Note: the agent must purchase electricity from the retail market in case of no energy storage in the energy pool.

Based on the description above, the problem can be transformed to optimum decision making with multiple coupled states and the Markov Decision Process (MDP) could be a practical choice.

The Markov Decision Process (MDP) provides a mathematical framework for modeling decision making in situations

where outcomes are partially random and partially under the control of the decision maker. The decision maker could take any action in available state and subsequently the process will move into a new state s'. As a result, the energy trading process could be modeled as the Markov Decision Process (MDP) as follows: State-of-Charge (SOC), retail electricity price and community electricity price in a certain time slot constitute the current state of the system. The decision to buy or sell energy, charge or discharge and the associated actual amount of energy need to be made and the system will enter another state in the next time slot.

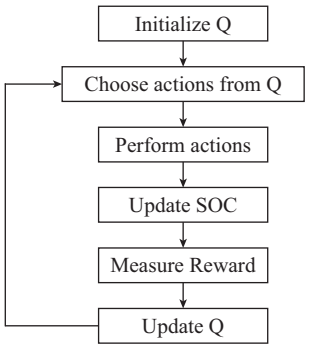


Fig. 2. Q learning brief algorithm.

III. Fuzzy Q-learning algorithm

A. The Reason for Using Q-learning

As described in Section II, the trading process could be modeled as a MDP problem. The Q-learning algorithm is one of the most effective ways to solve MDP problems because it is concerned with how agents ought to take actions in an environment; therefore, the cumulative reward could be maximized. Since Q-learning could be applied to model-free, ongoing tasks with excellent performance, it is advantageous for solving the proposed trading process quickly and accurately without actual modeling of market participants. As shown in Fig. 2, when used in the trading process (Fig. 2): it involves a set of states S namely SOC, retail price and community price; a set of actions per state A including all the combinations of buying or selling different amounts of electricity with the retail market or local energy pool and charging or discharging; and a Q-value table which is used to record Q-value Q(s,a) for different actions $a \in A$ when the agent is at state $s \in S$. The core of the Q-learning algorithm is the value iteration update, using the weighted average of the old value and the new information. And the agent could select the best set of actions with a maximum Q-value according to the Q-value table:

$$Q'(s,a) \leftarrow Q(s,a) + \alpha(R(s,a) + \gamma \max_{\mathrm{a}} Q(s',a) \qquad (9)$$

where $Q'(s,a)$ is the updated value, $Q(s,a)$ is the old value, $R(s,a)$ denotes the reward for the current state when taking action a and $\max_{\mathrm{a}} Q(s',a)$ means the optimal Q-value in next state x' which the agent could enter. The learning rate decides the extent of how the updated value overrides the old value. The discount factor determines the importance of future rewards. The agent would only consider the current rewards when $\gamma = 0$, while it will persist in long-term high rewards if $\gamma = 1$.

Finally, the agent of each market participant will record the stable action-state matrix. During the trading process, the market information is introduced into the agent and the agent would match the data to the status dimension of the matrix. Among all the actions with the highest matching degree, select the item with the highest Q value. In a nutshell, reinforcement learning helps to solve MDP problems with excellent performance.

B. Fuzzy Inference System for Battery

Q-learning algorithm could solve the Finite Markov Decision Process (FMDP) without the model of the environment and requiring adaptations. These advantages make Q-learning suitable for the smart community energy trading decision problem. Nevertheless, the battery model and price model are continuous and the continuous character will lead to an infinite state in Q-learning. It means Q-learning could never find the optimal points for this problem, while Fuzzy logic could obtain continuous variables from the discrete fuzzy set [24]. Therefore, a fuzzy Q-learning algorithm could be a rational choice.

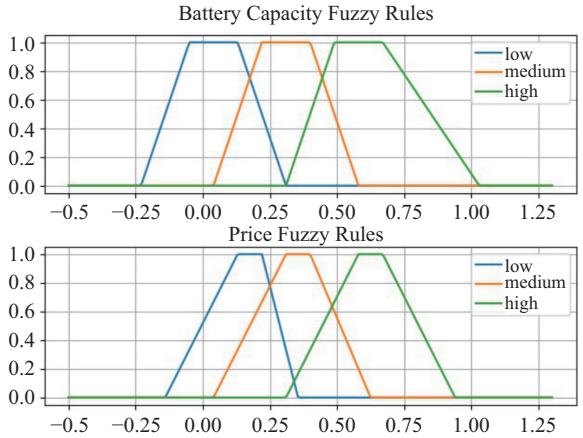


Fig. 3. Fuzzy Rules.

Fuzzy Logic is a multi-valued logic in which the truth value of the variables is translated into real numbers between 0 and 1. By using fuzzy inference, the values in the input vector can be translated to the corresponding values in the output based on a certain set of rules. Fuzzy rules are often defined as triangle or trapezoid-shaped curves and each value will have a slope where the value increases, a peak value of 1 (which can have a length of 0 or greater) and a slope where the value decreases. These rules are usually defined to express the extent of some attributes. That is to say, the infinite states of SOC and energy price that the Q-learning algorithm could not solve are now represented as finite states using the fuzzy rules (Fig. 3).

The fuzzy inference system can achieve a good approximation of the trading process and simultaneously make it feasible to use Q-learning in the continuous trading process.

C. Fuzzy Q-learning Algorithm

In the fuzzy Q-learning Algorithm, the set of states S is replaced by the AND combination of fuzzy output. A set of crisp inputs X were translated to fuzzy output defining the degree to which the agent is in a state. In addition, an exploration/exploitation algorithm was used to ensure the agent takes the action with the maximum Q-value in most situations and has the opportunity to take other actions that may have better future performance than the action with the maximum Q-value. Thus, the Fuzzy Q-learning algorithm has the following form:

1) Obtain the current state s.

2) Find optimal action combination for all fuzzy rules.

$$a = \begin{cases} \max Q_a(r_i, a) & \text{if } \ e < \varepsilon \\ \text{random}(a) & \text{if } \ e \geq \varepsilon, \end{cases} \tag{10}$$

where denotes the probability of exploitation, is the fuzzy rules.

3) Convolve the fuzzy output set $\omega(s)$ and plug it into the corresponding Q-value

$$Q(s,a) = \frac{\text{MIN}(\omega(s)) \cdot q(i,a)}{\text{sum}(\text{MIN}(\omega(s)))} \tag{11}$$

where $\text{MIN}(\omega(s))$ denotes the use of fuzzy operator to process the fuzzy output set $\omega(s)$, $q(i,a)$ denotes the q-value corresponding to the aggregated rule i for the selected action a.

4) Apply the selected action combination and obtain the new state s.

5) Calculate value function V(s',a) and Q-value variation

$$V(s',a) = \frac{\text{MIN}(\omega(s')) \max\limits_{a} q(i,a)}{\text{sum}(\text{MIN}(\omega(s')))} \tag{12}$$

$$\Delta Q = R(s,a,s') + \gamma \cdot V(s',a) - Q(s,a), \tag{13}$$

where $V(s',a)$ is the function to reach the maximum; Q-value for the new state s', $R(s,a,s')$ and γ have the same meaning with equation (9).

6) Update q-value according to:

$$q(i,a) \leftarrow q(i,a) + \alpha \cdot \Delta Q \cdot r(i,a), \tag{14}$$

where α denotes the learning rate, $r(i,a)$ is the defuzzificated truth value for the rule i with selected action a.

Algorithm 1: ALGORITHM FUZZY Q-LEARNING FOR ENERGY TRADING COMMUNITY

for iteration **in [1, max+1]:**
  for $s = 1, 2, 3$ **do:**
    compute fuzzy output set $\boldsymbol{\omega}(s)$
    end
  for fuzzy-value in $\boldsymbol{\omega}(s)$ **do:**
    convolution $\rightarrow r[i]$
    end
  for $i = 1, 2, \ldots, 27$ **do:**
    $Q \leftarrow Q + q(i \times a_s \times a_t) * r(i)$
    $V \leftarrow V + \max_{a_{si}, a_{ti}} q(i \times a_{si} \times a_{ti}) * r(i)$
    end
$Q \leftarrow Q / \sum_{i=1}^{27} r(i)$
$V \leftarrow V / \sum_{i=1}^{27} r(i)$
$\Delta Q = (s, a, s') + \gamma \cdot V - Q$
$q[i^*][a_s][a_t] \leftarrow q[i^*][a_s][a_t] + \alpha \cdot \Delta Q \cdot r(i^*)$
  for $i = 1, 2, \ldots, 27$ **do:**
   if $e < \varepsilon$:
     $a_s, a_t = \max_{a_s, a_t} q(i \times a_s \times a_t)$
     $s \leftarrow s'$
    end
  else:
  $a_s, a_t = \text{random}$
    end
end

D. Fuzzy MDP Framework for Energy Trading in the Community

According to the description about the community in Section II and previous sections, the process of energy trading for a smart home agent could be modeled as a MDP problem, which is given as follows:

- Community environment parameters change every 30 minutes, including retail price, community trade price and surplus-to-demand ratio.
- There are three state variables $\{s_1, s_2, s_3\}$ in the proposed system, where s_1 represents the battery capacity, s_2 represents the real-time retail price, s_3 represents the market price in the community and all variables are normalized in the range.
- There are two groups of actions for storage and electricity amount. For storage, $A_s = \{-1, -0.9, -0.8, \ldots, 0, \ldots, 0.8, 0.9, 1\}$ denotes the battery charge (+) / discharge (−) a_s of maximum charge/discharge rate at time t. For the community trading decision, $A_t = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ denotes the agent buy/sell a_t of the maximum electricity trade rate from the local energy pool.
- For each state, there are three fuzzy rules. In this paper, trapezoid-shaped curves are defined to judge the degree of capacity and price. Therefore, there will be $3 \times 3 \times 3 = 27$ kinds of rules' combination and the fuzzy system output $r(i)$ will be a 27-dimensional vector. Furthermore, the q-value table is a 3-demensional matrix with $q(i \times a_s \times a_t)$, where denotes the amount of battery action and denotes the amount of trade action.
- Reward function provides the agent with a reward for the selected action combination $\mathrm{a}_t = a_s \times a_t$ at state $\mathrm{s}_t = s_1 \times s_2 \times s_3$.

$$R(s,a,s') = -p_{\mathrm{n}} * \Delta\zeta(s) - \Delta p * \Delta\theta(s) \tag{15}$$

$$\Delta p = (p_{\mathrm{d}} + p_{\mathrm{b}} - p_{\mathrm{n}}) \tag{16}$$

$$\Delta\zeta(s) = \zeta(s) - \zeta(s') \tag{17}$$

$$\Delta\theta(s) = \theta(s) - \theta(s') \tag{18}$$

where p_{n} is the amount of trade electricity with the local energy pool ('+' denotes buy while '−' denotes sell); $\zeta(s)$ is the community market price at state s; $\theta(s)$ is the retail price at state s; p_{d} denotes the home energy demand; p_{b} denotes the amount of electricity the battery charge/discharge. And R(s,a,s') is normalized in the range [0,1].

- The local energy pool has to provide energy to non-intelligent users in the community besides the smart homes. As a result, the surplus-to-demand ratio will depend on the energy sold by the smart homes, the energy bought by all users and the solar energy.

IV. Case Study

A. System Initialization

In order to numerically examine the benefits that the fuzzy Q-learning algorithm could bring to a smart home user and the capability of consuming the surplus renewable energy, a community with five users and an energy pool was used as the test bed. The household ESS actions and the neighbor trade actions between smart home users and the energy pool were made every 30 mins, and the time frame spans the period from to the next.

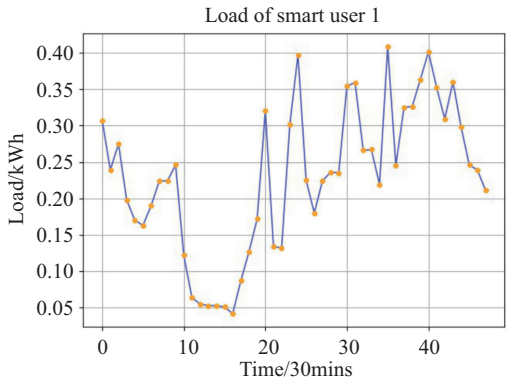


Fig. 4. Load data.

In this paper, all the data used to verify the correctness of the algorithm are real data, including demand data from the 5 users on the same street (e.g. smart user 1 in Fig. 4), the real-time retail price data in the British electricity retail market (Fig. 5) and the solar data for the solar photovoltaic (Fig. 6).

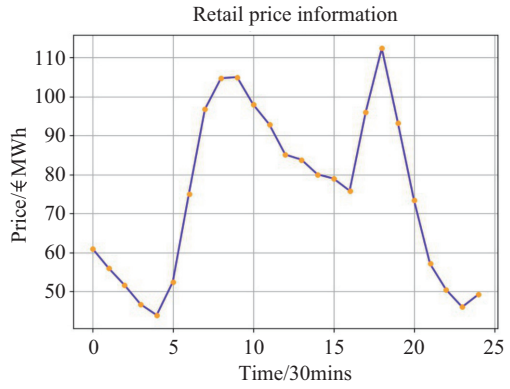


Fig. 5. Retail price data.

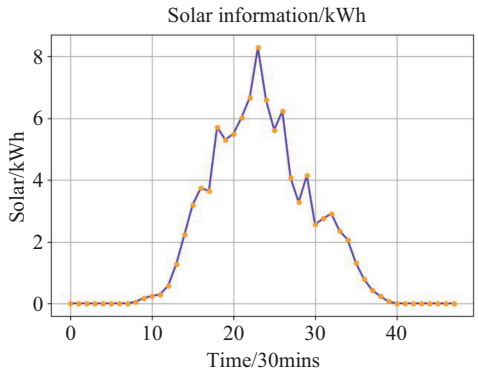


Fig. 6. Solar power generation data.

It assumes that the Tesla Powerwall was equipped in the smart home with capacity equal to 13.5 kWh and the maximum charge/discharge rate was equal to 5 kWh. Based on the total photovoltaic capacity installed in the community, the Feed-In-Tariff was set as 50 /MW.

For the fuzzy Q-learning algorithm proposed in chapter III, the exploration/exploitation rate was set to 0.99, $\gamma = 0.9$ and $\alpha = 0.1$. It means that the overall rewards were more important than the current rewards, which will facilitate the algorithm to generate more profit for users in the whole-time frame rather than the short-term immediate profit.

B. Case Study I: Only One Smart User in the Community

In this case, only one smart home user exists in the proposed community and all users' electricity consumption levels are similar. It suggests that only this user could sell redundant electricity to the local energy pool. This design is aimed to demonstrate the effect of a Fuzzy Q-learning Algorithm for improving smart home users' benefits.

Figure 7 shows the difference between the retail price and community price. Furthermore, the State-of-Charge (SOC) variation trend with price fluctuations was also shown in the figure. It is obvious that the community price is lower than the retail price most of the time because the redundant energy was sold to the energy pool and the solar energy will improve the surplus-to-demand ratio. However, in a small number of time slots, the energy pool price is higher than the retail price because the energy pool is guaranteed to be higher than the Feel-In-Tariff that aims to maintain the market participants' motivation. According to the algorithm proposed in chapter III, the agent will make optimal charge/discharge decisions for ESS, and the decision making is influenced by both the retail price and community price. For example, when $T = 16$ t, the retail price is relatively high and the community price is relatively low. According to the accumulated learning experience (that is q-value table), the agent determines that ESS supplying all of the household load is the best option with the maximum reward in the current state. In another example, when $T = 35$ t, both the retail price and community price reach the maximum in a day. As a result, the agent decides to discharge and sell energy to the local energy pool to earn more profit.

Figure 8 indicates the results of trade actions and storage actions at different times. When the ESS discharges, it ensures

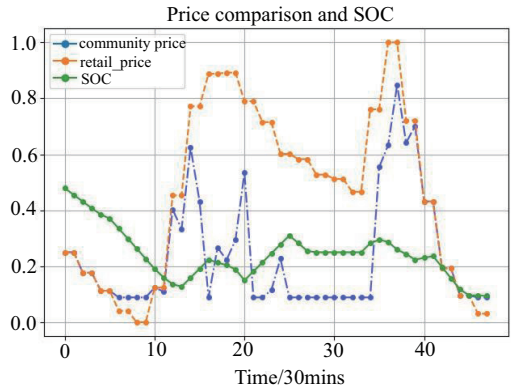


Fig. 7. The comparison between community price and retail price normalized in range.

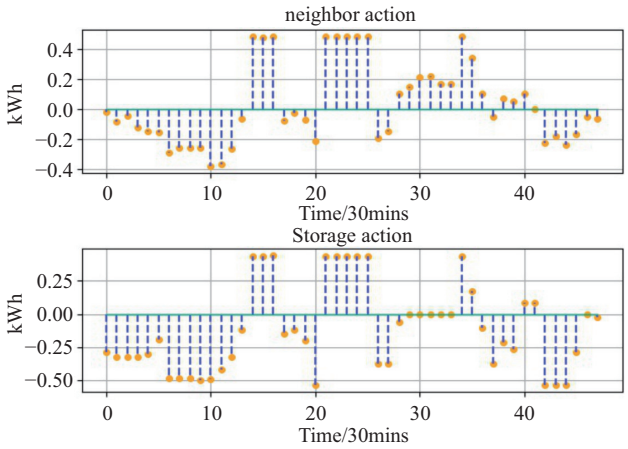


Fig. 8. Trade action and storage action at each time. + means charge and trade in while denotes discharge and trade out.

the uesr's demand is satisfied and then the remaining electricty would be sold to the energy pool. When the storage charges or the discharging electricty could not satisfy the user's demand, the agent purchases electricity from the energy pool. The time and quantity of the electricty trading was determined by the algorithm. For instance, from $T = 20$ t to $T = 25$ t, the solar energy is sufficient and the community price is relatively low. So the algorithm judges that the trading in and charging at this time could bring optimal overall rewards.

The total cost for each user is shown in Table I. Under the circumstances of a similar electricity consumption level, the smart user's cost for energy is significantly lower than the non-intelligent user. The results prove that the proposed algorithm could help users reduce their energy cost.

TABLE I
USERS' COST AMONG THE COMMUNITY

User ID	Cost
1 (Smart User)	−0.407
2 (non-intelligent)	−0.812
3 (non-intelligent)	−0.796
4 (non-intelligent)	−0.932
5 (non-intelligent)	−0.874

C. Case Study II: Two Smart Users in the Community

To verify the universality and the stability of the algorithm, a community with two smart users and three non-intelligent users was used. In the same way, all users' electricity consumption levels are similar. The initial SOC for smart user 1 is set to 0.5 and user 2 to 0.3.

Figure 9 demonstrates the change trend of the SOC. It can be observed that the algorithm makes similar decisions for two users in the same environment. To prove the stability of the algorithm, data from 0 to 2 am the next morning was provided. Through a day of charging and discharging, the state of charge could be maintained in an acceptable range at $T = 48$ t, which means the battery could store enough energy for the users and at the same time did not waste the surplus energy.

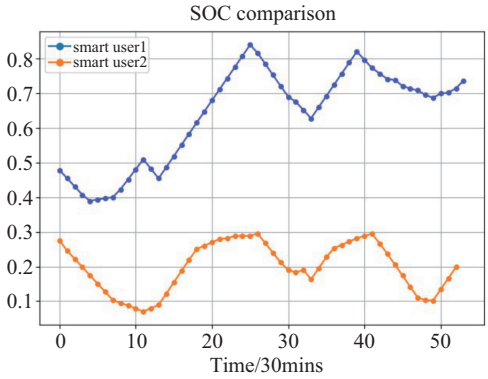


Fig. 9. The SOC comparison between two smart users.

Figures 10 and 11 show the trade actions and storage actions in a day from $T = 0$ to $T = 48$ t for the two smart users. Similar to what is discussed for Fig. 7, the proposed DSM approach could make beneficial decisions for the users. In addition, the trade actions taken by the two agents are similar, which further proves the universality of the algorithm. According to the results presented in TABLE II, the proposed DMS approach could help save costs for both smart users in the community. Therefore, the proposed DSM approach has solid universality and stability.

TABLE II
COST COMPARISON

User ID	Cost without algorithm	Cost with algorithm
1	−0.878	−0.613
2	−0.812	−0.635

D. Case Study III: One Smart User with Low Renewable Energy Penetration

One of the goals of using the Fuzzy Q-learning algorithm is to facilitate the consumption of renewable energy. Because the surplus-to-demand ratio depends primarily on the renewable energy, for example, the ratio is high when the solar energy is sufficient at noon. In order to prove the proposed algorithm is beneficial to optimize the consumption of renewable energy, the algorithm is run in the environment as Case I except that the penetration of renewable energy is also included. In this

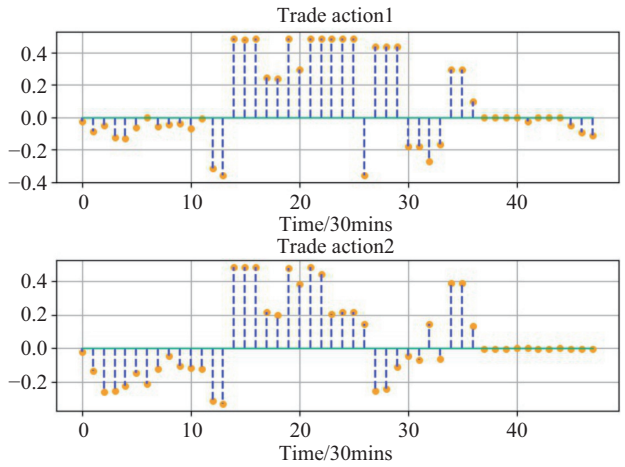


Fig. 10. Two users' trade action.

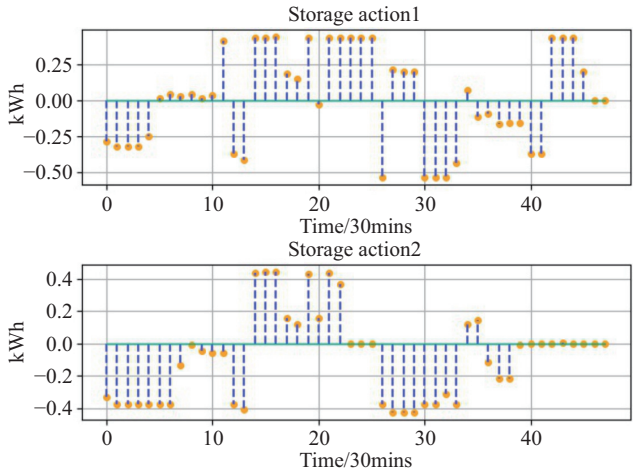


Fig. 11. Two users' storage action.

case, it is assumed that only 30% of solar energy is collected by the local energy pool, which could lead to an increase in the community price and reduce the participation desire for community energy trade.

Figure 12 shows the difference between the retail price and community price with low solar penetration. Compared with Fig. 7, the community price becomes higher because less solar energy was collected by the local energy pool. It will reduce the return ratio for the smart users' investment on their battery storage devices and less users would likely participate in the community energy trading. Furthermore, lots of renewable energy could be wasted because few users would buy electricity from the energy pool, which is completely opposite to the design. In addition, the electricity cost of the smart user is in this case which is significantly higher than that in case I.

Figure 13 demonstrates the result of smart agent trade actions with the local energy pool when the penetration level of solar energy is low. It is shown that from $T = 28\ t$ to $T = 40\ t$, there is no electricity trading between the energy pool and smart user. The main reason is the high community price caused by low solar energy penetration and users would

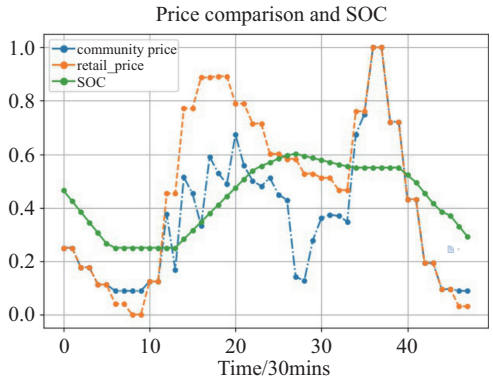


Fig. 12. The comparison between community price and retail price and the SOC with low solar penetration.

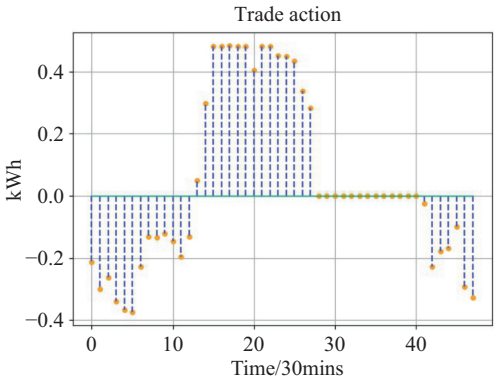


Fig. 13. The smart user trade action with low solar energy penetration in the community.

rather use the electricty in storage or in the retail market. All phenomenon observed above proves the correctness of the algorithm. The agent constantly improves the effectiveness of its energy management by experience learning and can identify the best option based on the existing q-value table. Though it is not as effective as the gloabal optimization solution, it could realize model-free and continuous online operation which could not be achieved by global optimization. In terms of the monetary benefit, the proposed algrithm can help smart users save considerable electricity cost and facilitate solar energy consumption.

In summary, when using the proposed Fuzzy Q-learning algorithm, high penetration of renewabale energy could bring more profits for users, which will consequently facilitate the development of renewable energy.

V. Conclusion

This paper proposes a smart energy community framework consisting of smart home users, non-intelligent users and a local energy pool that aims to facilitate the energy sharing among neighborhoods. The proposed community model allows neighborhoods to trade surplus energy to maximize the utilization of renewable energy. A pricing mechanism of the energy pool is also presented based on the demand-vs-surplus-ratio. To help the neighborhoods make the trading decisions, the fuzz Q-learning algorithm, a model free algorithm, is proposed for

the smart energy community. To evaluate the effectiveness of the proposed energy community framework and the algorithm, a series of numerical analysis have been performed under different scenarios. We discovered that the energy bills of the neighborhoods can be reduced in various scenarios. The main conclusions and results of this paper are summarized as follows:

1) A detailed model of a smart energy community with a local energy pool is proposed. With the proposed pricing mechanism for the energy pool, the energy pool price is effectively maintained between FIT and the retail market price. The results showed that such a pricing mechanism can help the pure energy consumers reduce their energy bills and increase the profits of energy prosumers.

2) The proposed Fuzz Q-learning algorithm can be applied to continuous and ongoing tasks such as energy trading in the community and this feature makes it feasible to constantly train the intelligent agents and improve the efficiency of its energy management strategies and decisions. Moreover, it is a model-free algorithm and optimal decisions can be made in real-time merely based on the q-value without complicated and tedious modeling of the community.

3) A smart community model that is capable of achieving P2P trading was presented, in which distributed household renewable energy generation was replaced by a centralized energy pool and the cost of equipment installation for users was eliminated. Users in the community could purchase electricity with lower prices from the local energy pool, and smart users could sell additional energy to the energy pool to generate extra income. Furthermore, the proposed smart community concept can lead to the transformation in the user role from consumer to prosumer. In this way, the users are able to participate in the electricity market, influencing the electricity price and achieving greater profits with the assistance of intelligent storage and control equipment.

The validation results demonstrated that the reinforcement learning is an effective way to solve the MDP problem in the DSM and the fuzzy inference system can provide a good approximation of a continuous process which makes it possible for an application of the Q-learning algorithm in a continuous problem. However, the approximate solutions might introduce unexpected error and instability. Hence, the future work will attempt to apply deep Q-learning and develop a subtilized model for the infinite MDP problem.

REFERENCES

[1] X. J. Li, D. Hui, and X. K. Lai, "Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 464–473, Apr. 2013.

[2] J. H. Yoon, R. Baldick, and A. Novoselac, "Dynamic demand response controller based on real-time retail price for residential buildings," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 121–129, Jan. 2014.

[3] D. Manna, S. K. Goswami, and P. K. Chattopadhyay, "Droop control for micro-grid operations including generation cost and demand side management," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 3, pp. 232–242, Sep. 2017.

[4] S. Y. Zhou, Z. Wu, J. N. Li, and X. P. Zhang, "Real-time energy control approach for smart home energy management system," *Electric Power Components and Systems*, vol. 42, no. 3–4, pp. 315–326, Feb. 2014.

[5] Z. Wu, X. P. Zhang, J. Brandt, S. Y. Zhou, and J. N. Li, "Three control approaches for optimized energy flow with home energy management system," *IEEE Power and Energy Technology Systems Journal*, vol. 2, no. 1, pp. 21–31, Mar. 2015.

[6] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 381–388, Aug. 2011.

[7] L. Gelazanskas and K. A. A. Gamage, "Demand side management in smart grid: A review and proposals for future direction," *Sustainable Cities and Society*, vol. 11, pp. 22–30, Feb. 2014.

[8] A. H. Mohsenian-Rad, V. W. S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[9] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-time demand response model," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 236–242, Dec. 2010.

[10] Z. M. Fadlullah, D. M. Quan, N. Kato, and I. Stojmenovic, "GTES: An optimized game-theoretic demand-side management scheme for smart grid," *IEEE Systems Journal*, vol. 8, no. 2, pp. 588–597, Jun. 2014.

[11] Y. Park and S. Kim, "Game theory-based bi-level pricing scheme for smart grid scheduling control algorithm," *Journal of Communications and Networks*, vol. 18, no. 3, pp. 484–492, Jun. 2016.

[12] Q. W. Wu, P. Wang, and L. Goel, "Direct load control (DLC) considering nodal interrupted energy assessment rate (NIEAR) in restructured power systems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1449–1456, Aug. 2010.

[13] C. Chen, J. H. Wang, and S. Kishore, "A distributed direct load control approach for large-scale residential demand response," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2219–2228, Sep. 2014.

[14] F. Zhang, R. de Dear, and C. Candido, "Thermal comfort during temperature cycles induced by direct load control strategies of peak electricity demand management," *Building and Environment*, vol. 103, pp. 9–20, Jun. 2016.

[15] K. Stenner, E. R. Frederiks, E. V. Hobman, and S. Cook, "Willingness to participate in direct load control: The role of consumer distrust," *Applied Energy*, vol. 189, pp. 76–88, Mar. 2017.

[16] R. de S?Ferreira, L. A. Barroso, P. R. Lino, M. M. Carvalho, and P. Valenzuela, "Time-of-use tariff design under uncertainty in price-elasticities of electricity demand: A stochastic optimization approach," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2285–2295, Dec. 2013.

[17] R. Li, Z. M. Wang, C. H. Gu, F. R. Li, and H. Wu, "A novel time-of-use tariff design based on Gaussian Mixture Model," *Applied Energy*, vol. 162, pp. 1530–1536, Jan. 2016.

[18] V. Hamidi, F. R. Li, L. Z. Yao, and M. Bazargan, "Domestic demand side management for increasing the value of wind," in *Proceedings of 2008 China International Conference on Electricity Distribution*, 2008, pp. 1–10.

[19] Y. Ozturk, D. Senthilkumar, S. Kumar, and G. Lee, "An intelligent home energy management system to improve demand response,?*IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 694–701, Jun. 2013.

[20] H. Aalami, G. R. Yousefi, and M. P. Moghadam, "Demand response model considering EDRP and TOU programs," in *Proceedings of 2008 IEEE/PES Transmission and Distribution Conference and Exposition*, 2008, pp. 1–6.

[21] R. Alvaro-Hermana, J. Fraile-Ardanuy, P. J. Zufiria, L. Knapen, and D. Janssens, "Peer to peer energy trading with electric vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 33–44, Jul. 2016.

[22] B. Celik, R. Roche, D. Bouquain, and A. Miraoui, "Coordinated neighborhood energy sharing using game theory and multi-agent systems," in *Proceedings of 2017 IEEE Manchester PowerTech*, 2017, pp. 1–6.

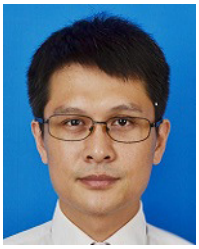
[23] C. Long, J. Z. Wu, C. H. Zhang, L. Thomas, M. Cheng, and N. Jenkins, "Peer-to-peer energy trading in a community microgrid," in *Proceedings of 2017 IEEE Power & Energy Society General Meeting*, Chicago, IL, 2017, pp. 1–5.

[24] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv: 1509.02971, 2015.

Suyang Zhou received the B.Eng. degree from Huazhong University of Science and Technology in 2009 and Ph.D. degree from University of Birmingham in 2015, both in Electrical Engineering. He is now a Lecturer with School of Electrical Engineering, Southeast University, Nanjing. Prior to joining Southeast University, he worked as KTP Associate at University of Leicester and Research and Development Engineer at Cellcare Technology Ltd between 2015 and 2016, and worked as Data Scientist at Power Networks Demonstration Centre (a joint research centre between University of Strathclyde and UK distribution network operators) between 2016 and 2017. His main research interests include integrated energy system, artificial intelligence in electrical domain and the demand side management.

Zijian Hu received the B.Eng. degree in Electrical Engineering from Southeast University, Nanjing, China, in 2018.

He is currently pursuing the Master degree at the School of Electrical Engineering, Southeast University. His research interests include the adoption of artificial intelligence on integrated energy system, and the control methodology for microgrid.



Wei Gu (M'06–SM'16) received the B.Eng. and Ph.D. degrees in Electrical Engineering from Southeast University, China, in 2001 and 2006, respectively. From 2009 to 2010, he was a Visiting Scholar with the Department of Electrical Engineering, Arizona State University, Tempe, AZ, USA. He is currently a Professor with the School of Electrical Engineering, Southeast University. His research interests include distributed generations and microgrids and active distribution networks.

Meng Jiang is an Assistant Professor in the Department of Computer Science and Engineering at the University of Notre Dame. He joined the faculty in 2017, after completing his Ph.D. in Computer Science and Technology at Tsinghua University in 2015 and postdoctoral training in Computer Science at the University of Illinois at Urbana-Champaign. His research focuses on machine learning for data-driven decision making, user behavior modeling, and information extraction.

Xiao-Ping Zhang (M'95–SM'06) received the B.Eng., M.Sc., and Ph.D. degrees in Electrical Engineering from Southeast University, China in 1988, 1990, 1993, respectively. He is currently Professor in Electrical Power systems and Director of Smart Grid of Birmingham Energy Institute at the University of Birmingham, U.K. Before joining the University of Birmingham, he was an Associate Professor in the School of Engineering at the University of Warwick, U.K. From 1998 to 1999, he was visiting UMIST. From 1999 to 2000, he was an Alexander-von-Humboldt Research Fellow with the University of Dortmund, Germany. He worked at China Electric Power Research Institute on EMS/DMS advanced application software research and development between 1993 and 1998. He is co-author of the monograph Flexible AC Transmission Systems: Modeling and Control (New York: Springer, 2006 and 2012). Prof Zhang is an editor of the *IEEE Transactions on Smart Grid*. Internationally, Professor Zhang pioneered the concept of "Global Power & Energy Internet" and "Energy Union."