Multiagent Reinforcement Learning for Community Energy Management to Mitigate Peak Rebounds Under Renewable Energy Uncertainty

Bo-Chen Lai, Wei-Yu Chiu, *Member, IEEE*, and Yuan-Po Tsai

***Abstract*—Price-based demand response (DR) can aid power grid management, but an uncoordinated response may lead to peak rebounds during low-price periods. This article proposes a community energy management system based on multiagent reinforcement learning. The scheme consists of a community aggregator that optimizes the total community electricity cost for multiple residential users. A home requires energy management for home appliances, electric vehicles, energy storage systems, and renewable energy generation. The appliance scheduling problem is decomposed into smaller sequential decision problems that are easier to solve. Renewable generation is predicted and used to mitigate the influence of energy generation uncertainty. As indicated in numerical analyses, the proposed approach can handle the uncertainty in renewable energy and leads to more economical energy usage relative to existing energy management methods. The method outperforms conventional algorithms, such as centralized mixed-integer nonlinear programming and genetic algorithm-based optimization, in terms of mitigating peak rebounds and addressing the uncertainty of renewable energy generation.**

***Index Terms*—Appliance scheduling, energy management system, game theory, multiagent reinforcement learning, neural network, peak rebound, renewable energy sources.**

NOMENCLATURE

l_n	Power consumption information of residential user n
p^{ag}	Power imported from the utility to the community
$p^{\text{non}}_{n,t}$	Power consumption of non-shiftable appliances in active residence n at time t
$D^{\text{non}}_{n,t}$	Power demand of non-shiftable appliances in residence n at time t
T	Observation window
$u^{\text{shift}}_{n,i,t}$	Indicator function of shiftable appliance i in residence n at time t
$T^{\text{shift}}_{n,i,\text{start}}, T^{\text{shift}}_{n,i,\text{end}}$	Start and end times of shiftable appliance i in residence n, respectively
$w^{\text{shift}}_{n,i}$	Nominal total working hour of shiftable appliance i in residence n
$D^{\text{shift}}_{n,i,t}$	Power demand of shiftable appliance i in residence n
$p^{\text{shift}}_{n,i,t}$	Power consumption of shiftable appliance i in residence n at time t
$g^{\text{shift}}_{n,i,t}$	User dissatisfaction cost of shiftable appliance i in residence n at time t
$\beta^{\text{shift}}_{n,i}$	Dissatisfaction coefficient of shiftable appliance i in residence n
$p^{\text{con}}_{n,j,t}$	Power consumption of controllable appliance j in residence n at time t
$D^{\text{con}}_{n,j,\min}, D^{\text{con}}_{n,j,\max}$	Minimum and maximum power demand of controllable appliance j in residence n, respectively
$T^{\text{con}}_{n,j,\text{start}}, T^{\text{con}}_{n,j,\text{end}}$	Start and end times for controllable appliance j in residence n, respectively
$g^{\text{con}}_{n,j,t}$	User dissatisfaction cost of controllable appliance j in residence n at time t
$\beta^{\text{con}}_{n,j}$	Dissatisfaction coefficient of controllable appliance j in residence n
$p^{\text{EV}}_{n,t}$	Power consumption of charging an EV in residence n at time t
$g^{\text{EV}}_{n,t}$	Charging anxiety function for an EV
β^{EV}_{n}	Charging anxiety coefficient
$D^{\text{EV}}_{n,\min}, D^{\text{EV}}_{n,\max}$	Minimum and maximum charging demand of an EV in residence n
$T^{\text{EV}}_{n,\text{start}}, T^{\text{EV}}_{n,\text{end}}$	Start and end times for an EV charging event in residence n, respectively
$B_{n,t}, p^{\text{B}}_{n,t}$	Energy level and charging/discharging power of an ESS in residence n at time t
η_n	Charging and discharging efficiency of the ESS in residence n
$B^{\min}_n, B^{\max}_n$	Minimum energy level and maximum capacity of the ESS in residence n
$p^{\min}_{n,t}, p^{\max}_{n,t}$	Discharge and charge limits of the ESS in residence n
$p^{\text{PV}}_{n,t}$	Solar power in residence n at time t
f_n	Electricity cost of residential user n
λ_t	Electricity price in time slot t
Δt	Duration of a time slot
C^{deg}_n	Degradation cost of the ESS
g_n	Dissatisfaction cost of residential user n

Manuscript received September 13, 2021; revised January 27, 2022; accepted February 14, 2022. Date of publication March 21, 2022; date of current version May 26, 2022. This work was supported by the Ministry of Science and Technology of Taiwan under Grant MOST 110-2221-E-007-097-MY2. *(Corresponding author: Wei-Yu Chiu.)*

The authors are with MOCaRL Lab, Department of Electrical Engineering, National Tsing Hua University, Hsinchu 300044, Taiwan (e-mail: nicklai.bcl@gmail.com; chiuweiyu@gmail.com; ylsh910459@gmail.com).

Digital Object Identifier 10.1109/TETCI.2022.3157026

U_n	Cost function of residential user n
w_n	Weight that reflects the desired balance between electricity cost and dissatisfaction cost of residential user n
$\boldsymbol{P}_n$	Power control vector for all appliances and the ESS in residence n
B_t, p_t^{B}	Energy level and charging/discharging power of an ESS at CA
η	Charging/discharging efficiency of the ESS at CA
$B^{\min}, B^{\max}$	Minimum energy level and maximum capacity of the ESS at CA
$p_t^{\min}, p_t^{\max}$	Discharge and charge limits of the ESS at CA, respectively
p_t^{CA}	Amount of power imported from the grid to the whole community
p_t^{PV}	Solar power at CA at time t
$\mathcal{N}, \mathcal{M}$	Sets of active and passive residential users, respectively
$L_t^{\min}, L_t^{\max}$	Lower and upper bounds of p_t^{CA}
U	Cost function of the CA
C^{deg}	Degradation cost of the ESS at CA
τ	Overprice that forces the CA to meet the constraints induced by $L_t^{\min}$ and $L_t^{\max}$
$\boldsymbol{p}^{\mathrm{B}}$	Charging/discharging control vector at CA
C_t	Generation cost on the supply side
e_t, f_t	Electricity price coefficients
$\mathcal{C}$	Set of all followers
$\boldsymbol{l}_n$	Power demand vector
Γ	Non-cooperative Stackelberg game (NSG) strategy
w	weather forecast information
$LTSM(\cdot)$	LSTM prediction model of renewable energy generation
$\mathcal{S}, s_t$	State space and state at time t, respectively
$\mathcal{A}, a_t$	Action space and action at time t, respectively
r_t	Cost.
π	Policy that maps a state to an action
π_*	Optimal policy.
q_π	Action-value function
γ	Discount factor
$q_*(s_t, a_t)$	Optimal action-value function
$Q(S_t, A_t)$	Q-value at state-action pair (S_t, A_t)
$\boldsymbol{d}_h$	Demand load profile of upcoming time slots starting from time h and ending at time T
$\boldsymbol{p}_h^{\mathrm{PV}}$	Forecast PV generation at CA
$Q_{n,k}$	Q-value of appliance k in residence n
$\boldsymbol{l}_{n,h}$	Power demand of the selected agent n at time h
$\boldsymbol{l}_{-n,h}$	Power demand of other unselected agents at time h

I. Introduction

BECAUSE of the continual changes in the electricity market with respect to, for example, electricity prices and user energy consumption, an energy management system (EMS) that adaptively optimizes generation or power transmission is required [1]. Moreover, recent advances in information transmission and smart metering technologies have led to an increased focus on demand response (DR) strategies that improve grid efficiency and reliability by adjusting flexible loads on the demand side [2]. A well-designed DR program can aid power grid management by balancing electricity supply and demand, facilitating the use of renewable energy sources, and reducing fossil fuel consumption [3].

With the increasing use of smart home appliances, electric vehicles (EVs), and energy storage systems (ESS), home energy management systems (HEMS) based on price-based DR provide new opportunities to achieve energy efficiency through efficient energy scheduling without any compromise to user satisfaction. HEMS can reduce both the electricity cost and peak-to-average ratio by shifting some appliance operations to a period when electricity prices are low [4]. Additionally, user satisfaction and comfort levels in relation to thermoelectrical loads should be considered [5], and HEMS can help balance between cost and user satisfaction.

However, in a typical DR program, residential users may receive the same price signal from the utility, thereby raising the risk that many users would operate their appliances during the same low-price period. This effect is referred to as a peak rebound [6]–[8] and a systemwide DR management mechanism is required to coordinate residential user and community levels. Two coordination structures to mitigate peak rebound have been examined, a centralized structure and distributed coordination structure, which are classified according to the underlying communication and control architecture [9].

A centralized structure has a central operator managing the electricity use of all smart homes, with direct access to information relating to the electric appliances of all end users [10], [11]. Although some studies have demonstrated that the centralized approach is optimal for electrical energy use, one major drawback is the computational burden during optimization, especially because of the large number of residential user assets that must be controlled [12]. Furthermore, the requirement for detailed information on end users may raise some privacy concerns [13].

By contrast, a distributed structure can distribute computation to several subsystems to mitigate the high computational burden and privacy concerns [14]. A distributed structure allows end users to schedule their loads individually while communicating with a central entity to obtain information about neighboring electricity profiles. Community energy management can be decomposed into a two-level optimization problem, in which the upper level seeks to flatten the system load profile and the lower level minimizes individual residential users' energy costs [15], [16].

A popular approach used for the distributed structure is game theory [8], [17], [18]. Zhu *et al.* [19] proposed a noncooperative game based on mixed-integer programming to schedule consumption plans to minimize energy costs for several residential consumers, but they only considered nonshiftable and shiftable appliances in their scenario. Li *et al.* [20] proposed a distributed algorithm for shiftable appliances to minimize energy costs.

Rajasekhar *et al.* [21] examined an energy scheduling problem in a residential community with an aggregator. In that study, EVs; batteries; and critical, controllable, and shiftable loads were considered. The results indicated that the home appliance scheduling problem became difficult to solve the more home appliances there are. Genetic algorithms and the Stackelberg game were applied to optimize users' electricity costs and satisfaction.

Although previous studies considering distributed coordination have addressed the peak rebound problem, they have implicitly assumed that the information on renewable energy generation is accurate. Based on this assumption, optimization methods can perform excellently. Given the lack of complete environmental information, such as the uncertainty in electricity prices or photovoltaic (PV) generation, reinforcement learning (RL) can outperform conventional optimization methods [22]–[24]. For example, Remani *et al.* [23] presented an innovative RL-based model for residential shiftable load scheduling based on uncertain renewable energy. To consider the diversity of home appliances, Xu *et al.* [24] proposed multiagent RL (MARL)-based HEMS that considered electricity prices and renewable energy uncertainty but did not include an ESS. Although existing RL-based energy management methods have addressed the problem of uncertainty in renewable energy, the focus has primarily been on a single residence, potentially leading to peak rebound.

Advances in community energy management in a distributed structure have two weaknesses. First, most studies have formulated the scheduling of home appliances as a problem in which the computation time dramatically increases with an increasing number of appliances. Second, existing approaches addressing peak rebound have not considered the uncertainty of renewable energy generation, and approaches addressing uncertainty have focused on a single residence instead of a whole community.

To fill this research gap, we propose a MARL algorithm for community energy management to mitigate peak rebound influenced by uncertain renewable energy generation. A leader–follower Stackelberg game is formulated in each time step. In the game, a community aggregator (CA) serving as the leader forecasts future renewable energy generation, which can be achieved using a method known as weather-based long short-term memory (LSTM) [25]. The CA optimizes the ESS scheduling while updating a Q-table and initializes a community load profile for all residential users. Residential users acting as followers predict their own renewable energy generation.

We decompose the scheduling problem of home appliances into a sequential scheduling problem for a single appliance. In each home appliance scheduling problem, an appliance in a residence is an agent, which has its own Q-table; this entails multiple agents in the residence. Once an agent has updated the Q-table in a planning phase, it transfers information about the available renewable energy to the next appliance agent for scheduling. After all the appliances have been scheduled, the residential user notifies the CA of their load profile. This interaction between the CA and residential user continues in that time slot until a near Stackelberg equilibrium is reached. In the next time slot, a new Stackelberg game is formulated, and the appliance agents can inherit Q-tables obtained from the previous time slot. Simultaneously, weather information is used to predict renewable generation for the appliance agents to adjust their scheduling to mitigate the influence of uncertainty.

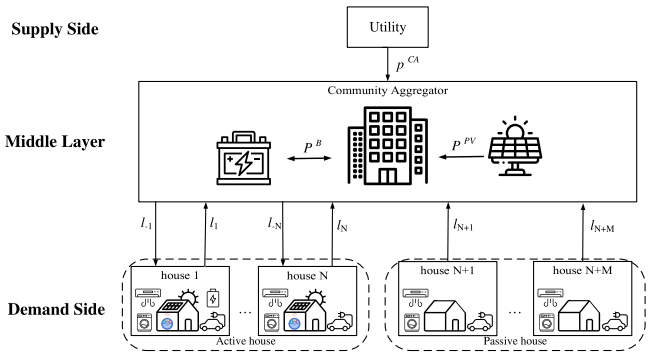

Fig. 1. Bi-level DR structure.

This study's main contributions are as follows. To mitigate the influence of uncertain renewable energy generation, the proposed community energy management method incorporates future renewable energy generation prediction into energy scheduling. To mitigate peak rebounds for community energy management, we formulated the problem of appliance scheduling using a game-theoretic approach, with MARL used to solve the game. We decomposed the scheduling problem of home appliances into smaller sequential decision problems: individual energy scheduling problems for shiftable appliances, controllable appliances, EVs, and ESS. The proposed MARL can efficiently use renewable energy because of the information transfer between agents. The proposed method was compared with existing learning-based and optimization-based methods for community energy management. Compared with a single-agent RL-based method, our method reduced the peak load and average cost by 30.8% and 9.68%, respectively. Compared with centralized mixed-integer nonlinear programming (MINLP), our method reduced the average residential cost by 2.7%, the standard deviation of the cost by 56%, and the peak load at the CA by 1.72% while handling the uncertainty in renewable energy generation.

The rest of this article is organized as follows. The system description, mathematical models relating to home appliances and ESS, and problem formulation are presented in Section II. The proposed community energy management method is detailed in Section III. The simulation results are presented in Section IV. Finally, the conclusion is presented in Section V.

II. System Models

This section presents mathematical models representing residential communities; the models include the supply side, middle layer, and demand side and comprise N active residential users and M passive residential users (Fig. 1) [16], [20], [23]. Each active residential user has an HEMS, renewable energy source (e.g., solar panels), and ESS. Active residential users can schedule the operating periods of household appliances to minimize their electricity spending with acceptable user satisfaction in

response to time-varying electricity prices. Passive residential users neither optimize the appliance schedule nor respond to price changes. In this scheme, residential user n provides the CA with their power consumption information l_n in the middle layer. The CA is equipped with an ESS [21], receives all power consumption profiles, and minimizes community electricity spending. The CA can import power p^{CA} from the utility to balance supply and demand.

A. Residence Models

The HEMS of residential users can forecast future renewable energy and optimize the scheduling of appliances to reduce electricity bills. Appliances in a smart home can be classified into three types [22]: nonshiftable appliances, shiftable appliances, and controllable appliances. In addition, each residence is assumed to be equipped with an EV, ESS, and PV system [21].

1) Nonshiftable Appliances: Nonshiftable appliances, such as refrigerators and cooking appliances, are commonly used [26]. These appliances are inflexible and cannot be scheduled. We denote $p_{n,t}^{\text{non}}$ kW as the power consumption of nonshiftable appliances in an active residence n in time slot t. The total power consumption of the nonshiftable appliances is equal to the power demand $D_{n,t}^{\text{non}}$ and can be expressed as

$$p_{n,t}^{\text{non}} = D_{n,t}^{\text{non}}, \quad t = 1, 2, \ldots, T \tag{1}$$

where T is the observation window.

2) Shiftable Appliances: Shiftable appliances, such as dishwashers and washing machines, can be shifted from a high-price period to a low-price period to reduce electricity costs [26]. Unlike nonshiftable appliances, shiftable appliances have two available actions to choose from: "on" and "off," coded as 1 and 0, respectively. If residence n has I_n shiftable appliances, let $u_{n,i,t}^{\text{shift}}$ denote the indicator function:

$$u_{n,i,t}^{\text{shift}} \in \{0, 1\} \quad \forall t \in [T_{n,i,\text{start}}^{\text{shift}}, T_{n,i,\text{end}}^{\text{shift}}], \quad i = 1, 2, \ldots, I_n \tag{2}$$

where $T_{n,i,\text{start}}^{\text{shift}}$ and $T_{n,i,\text{end}}^{\text{shift}}$ represent the start and end times of shiftable appliance i, respectively.

A shiftable appliance must complete its nominal total working hour $w_{n,i}^{\text{shift}}$ in $[T_{n,i,\text{start}}^{\text{shift}}, T_{n,i,\text{end}}^{\text{shift}}]$; thus, the constraint on the indicator function can be expressed as

$$\sum_{t=T_{n,i,\text{start}}^{\text{shift}}}^{T_{n,i,\text{end}}^{\text{shift}}} u_{n,i,t}^{\text{shift}} = w_{n,i}^{\text{shift}}. \tag{3}$$

Power consumption $p_{n,i,t}^{\text{shift}}$ is equal to power demand $D_{n,i,t}^{\text{shift}}$ kW controlled by the indicator function:

$$p_{n,i,t}^{\text{shift}} = u_{n,i,t}^{\text{shift}} D_{n,i,t}^{\text{shift}}. \tag{4}$$

Although energy costs can be reduced by shifting the power demand of the shiftable appliances to time slots with low electricity prices, this practice can cause user dissatisfaction by delaying the expected schedule. The dissatisfaction cost of shiftable appliances can be modeled as [23]

$$g_{n,i,t}^{\text{shift}} = \beta_{n,i}^{\text{shift}}(1 - u_{n,i,t}^{\text{shift}}) D_{n,i,t}^{\text{shift}} \quad \forall t \in [T_{n,i,\text{start}}^{\text{shift}}, T_{n,i,\text{end}}^{\text{shift}}] \tag{5}$$

where $\beta_{n,i}^{\text{shift}}$ is a dissatisfaction coefficient of appliance i in residence n.

3) Controllable Appliances: Controllable appliances can operate flexibly in a predefined power range, such as in air conditioners and water heaters. Let $p_{n,j,t}^{\text{con}}$ denote the power consumption of controllable appliance j. If residence n has J_n controllable appliances, the following constraints are imposed on $p_{n,j,t}^{\text{con}}$:

$$D_{n,j,\min}^{\text{con}} \le p_{n,j,t}^{\text{con}} \le D_{n,j,\max}^{\text{con}}, \quad j = 1, 2, \ldots, J_n \tag{6}$$

where $D_{n,j,\min}^{\text{con}}$ and $D_{n,j,\max}^{\text{con}}$ represent the minimum and maximum power demand, respectively.

Although controllable appliances can lower household electricity spending by decreasing the power consumption, the reduction in power may cause user dissatisfaction. The user dissatisfaction cost of controllable appliance j can be modeled as [27]

$$g_{n,j,t}^{\text{con}} = \beta_{n,j}^{\text{con}}(p_{n,j,t}^{\text{con}} - D_{n,j,\max}^{\text{con}})^2 \quad \forall t \in [T_{n,j,\text{start}}^{\text{con}}, T_{n,j,\text{end}}^{\text{con}}] \tag{7}$$

where $T_{n,j,\text{start}}^{\text{con}}$ and $T_{n,j,\text{end}}^{\text{con}}$ denote the start and end time slots, and $\beta_{n,j}^{\text{con}}$ is a dissatisfaction coefficient of controllable appliance j.

4) Electric Vehicles: Charging an EV introduces a controllable load. Let $p_{n,t}^{\text{EV}}$ denote the power consumption of charging an EV in residence n at time t. The following constraints are imposed on $p_{n,t}^{\text{EV}}$

$$D_{n,\min}^{\text{EV}} \le p_{n,t}^{\text{EV}} \le D_{n,\max}^{\text{EV}} \quad \forall t \in [T_{n,\text{start}}^{\text{EV}}, T_{n,\text{end}}^{\text{EV}}] \tag{8}$$

where $D_{n,\min}^{\text{EV}}$ and $D_{n,\max}^{\text{EV}}$ represent the minimum and maximum charging demand, respectively.

A "charging anxiety" function $g_{n,t}^{\text{EV}}$ of an EV represents the fear of having insufficient energy to charge an empty EV battery; the anxiety function is defined as [24]

$$g_{n,t}^{\text{EV}} = \beta_n^{\text{EV}}(p_{n,t}^{\text{EV}} - D_{n,\max}^{\text{EV}})^2 \quad \forall t \in [T_{n,\text{start}}^{\text{EV}}, T_{n,\text{end}}^{\text{EV}}] \tag{9}$$

where $T_{n,\text{start}}^{\text{EV}}$ and $T_{n,\text{end}}^{\text{EV}}$ are the start and end times for EV charging, respectively, and β_n^{EV} represents the charging anxiety coefficient.

5) Energy Storage Systems: Some residential users are assumed to be equipped with an ESS to store surplus renewable energy for future use, using batteries. Let $B_{n,t}$ and $p_{n,t}^{\text{B}}$ denote the energy level of the ESS and charging and discharging power, respectively. The storage dynamic can be expressed as [21]

$$B_{n,t+1} = B_{n,t} + \eta_n p_{n,t}^{\text{B}} \Delta t \tag{10}$$

where η_n represents the charging and discharging efficiency of the battery.

The constraints on the ESS are

$$B_n^{\min} \le B_{n,t} \le B_n^{\max} \tag{11}$$

where $B_n^{\min}$ and $B_n^{\max}$ are the minimum energy level and maximum storage capacity, respectively. The constraints on the charging and discharging power of the ESS are

$$p_{n,t}^{\min} \le p_{n,t}^{\text{B}} \le p_{n,t}^{\max} \tag{12}$$

where $p_{n,t}^{\min}$ and $p_{n,t}^{\max}$ are the discharging and charging limits of the ESS, respectively.

Some residences have solar panels on their roofs that produce renewable power $p_{n,t}^{\text{PV}}$. The power consumption $l_{n,t}$ of residential user n can be expressed as

$$l_{n,t} = \max\{p_{n,t}^{\text{non}} + \sum_{i=1}^{I_n} p_{n,i,t}^{\text{shift}} + \sum_{j=1}^{J_n} p_{n,j,t}^{\text{con}} + p_{n,t}^{\text{EV}} - p_{n,t}^{\text{PV}} + p_{n,t}^{\text{B}}, 0\}. \tag{13}$$

The electricity cost of residential user n, denoted as f_n, can be expressed as

$$f_n = \sum_{t=1}^{T} \lambda_t l_{n,t} \Delta t + C_n^{\text{deg}} |p_{n,t}^{\text{B}}| \tag{14}$$

where λ_t is the electricity price in time slot t, Δt is the duration of a time slot, and C_n^{deg} is the degradation cost of the battery. The dissatisfaction cost g_n of residential user n is

$$g_n = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{I_n} g_{n,i,t}^{\text{shift}} + \sum_{j=1}^{J_n} g_{n,j,t}^{\text{con}} + g_{n,t}^{\text{EV}} \right\}. \tag{15}$$

The cost function U_n of residential user n is defined as

$$U_n = w_n f_n + (1 - w_n) g_n \tag{16}$$

where $w_n \in (0, 1)$ represents the residential user's weighting, reflecting the desired balance between the electricity cost and dissatisfaction cost. Let

$$\boldsymbol{P}_n = \begin{bmatrix} \boldsymbol{u}_{n,1}^{\text{shift}} \\ \vdots \\ \boldsymbol{u}_{n,I_n}^{\text{shift}} \\ \boldsymbol{p}_{n,1}^{\text{con}} \\ \vdots \\ \boldsymbol{p}_{n,J_n}^{\text{con}} \\ \boldsymbol{p}_n^{\text{EV}} \\ \boldsymbol{p}_n^{\text{B}} \end{bmatrix} = \begin{bmatrix} u_{n,1,1}^{\text{shift}} & u_{n,1,2}^{\text{shift}} & \cdots & u_{n,1,T}^{\text{shift}} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n,I_n,1}^{\text{shift}} & u_{n,I_n,2}^{\text{shift}} & \cdots & u_{n,I_n,T}^{\text{shift}} \\ p_{n,1,1}^{\text{con}} & p_{n,1,2}^{\text{con}} & \cdots & p_{n,1,T}^{\text{con}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,J_n,1}^{\text{con}} & p_{n,J_n,2}^{\text{con}} & \cdots & p_{n,J_n,T}^{\text{con}} \\ p_{n,1}^{\text{EV}} & p_{n,2}^{\text{EV}} & \cdots & p_{n,T}^{\text{EV}} \\ p_{n,1}^{\text{B}} & p_{n,2}^{\text{B}} & \cdots & p_{n,T}^{\text{B}} \end{bmatrix} \tag{17}$$

be the power control vector for the appliances and battery. The goal of residential user n can be described as

$$\min_{\boldsymbol{P}_n} U_n \text{ subject to } (2)\ (3), (6),\ (8),\ (10),\ (11), \text{and } (12). \tag{18}$$

B. Aggregator Model

The CA is assumed to possess a battery and PV system and acts as a broker between the residential users and utility [21]. The goal of the CA is to minimize costs. To achieve this goal, its battery can be used to perform load shifting. Let B_t and p_t^{B} denote the energy level and charging and discharging power of an ESS at time t in the CA, respectively. The community storage dynamics can be expressed as [21]

$$B_{t+1} = B_t + \eta p_t^{\text{B}} \Delta t \tag{19}$$

where η represents the charging and discharging efficiency of the CA's ESS.

The constraints on the community storage system are

$$B^{\min} \le B_t \le B^{\max} \tag{20}$$

where $B^{\min}$ and $B^{\max}$ are the minimum energy level and maximum storage capacity in the CA, respectively. The constraints on the charging and discharging power of the battery are

$$p_t^{\min} \le p_t^{\text{B}} \le p_t^{\max} \tag{21}$$

where $p_t^{\min}$ and $p_t^{\max}$ are the discharging and charging limits of the ESS, respectively.

The amount of power for the whole community imported from the grid is

$$p_t^{\text{CA}} = \sum_{n \in \mathcal{N} \cup \mathcal{M}} l_{n,t} + p_t^{\text{B}} - p_t^{\text{PV}} \tag{22}$$

where p_t^{PV} represents renewable power and $\mathcal{N}$ and $\mathcal{M}$ are the sets of active and passive residential users, respectively. The imported power is further constrained by [28]

$$L_t^{\min} \le p_t^{\text{CA}} \le L_t^{\max}. \tag{23}$$

An aggregate power demand lower than $L_t^{\min}$ may result in additional costs if base load power plants are turned off; $L_t^{\max}$ is the upper boundary at which the aggregate demand without outage is satisfied.

The cost function of the CA, denoted by U, can be expressed as

$$U = \sum_{t=1}^{T} \{\lambda_t p_t^{\text{CA}} \Delta t + C^{\text{deg}} |p_t^{\text{CA}}| + \tau \max(0, L_t^{\min} - p_t^{\text{CA}}, p_t^{\text{CA}} - L_t^{\max})\} \tag{24}$$

where C^{deg} represents the degradation cost of the ESS and τ represents the overprice that forces the CA to satisfy the constraints in (23).

The electricity price per unit λ_t for the demand side can be modeled as [17], [19], [21]

$$\lambda_t = \frac{C_t}{p_t^{\text{CA}}} = e_t p_t^{\text{CA}} + f_t, \quad t = 1, 2, \ldots, T \tag{25}$$

where C_t represents the generation cost on the supply side, and e_t and f_t are price constants. The cost function C_t is a quadratic function of p_t^{CA}, which reflects a common assumption that the cost increases quadratically with power consumption. The model has been extensively used in research because it resembles a physical system but is nonetheless simple enough to be analyzed.

Let

$$\boldsymbol{p}^{\text{B}} = \begin{bmatrix} p_1^{\text{B}} & p_2^{\text{B}} & \cdots & p_T^{\text{B}} \end{bmatrix} \tag{26}$$

be the battery charging and discharging control vector of the CA. The goal of the CA can be achieved by solving

$$\min_{\boldsymbol{p}^{\text{B}}} U \text{ subject to } (19),\ (20),\ (21) \text{ and } (22). \tag{27}$$

III. Proposed Method for Distributed Community Energy Management

Cost functions of residential users and the CA are to be minimized. The associated solutions will affect each other and these problems should not be solved centrally due to privacy issues. These problems can form a non-cooperative Stackelberg (NSG) game. The CA plays the role of a leader in the game and the residential users act as followers [21]. We denote the set of all followers as $\mathcal{C} = \mathcal{N} \cup \mathcal{M}$.

In a Stackelberg game, the leader will make its decision in consideration of the best responses for the followers. The CA determines its battery dispatch profile $\boldsymbol{p}^{\mathrm{B}}$ to minimize the cost function U in (27). All residential users make their optimal decisions of power demand vector $\boldsymbol{l}_n = [l_{n,1} \quad l_{n,2} \cdots l_{n,T}]$ to minimize their cost function U_n in (18) after being notified of the leader's decision. The NSG strategy Γ can be formally defined by the following strategic form:

$$\Gamma = \{\mathcal{C} \cup \{\mathrm{CA}\}, \{\boldsymbol{l}_n\}_{n\in\mathcal{C}}, \{U_n\}_{n\in\mathcal{C}}, \boldsymbol{p}^{\mathrm{B}}, U\}. \tag{28}$$

The community energy scheduling problem is thus formulated as a Stackelberg game, and the solution is the Stackelberg equilibrium or near Stackelberg equilibrium in which the leader finds its optimal storage dispatch profile under the followers' equilibrium state, corresponding with the optimal power demand. To be specific, consider the NSG strategy Γ defined in (28) where U_n and U are determined by solving (18) and (27), respectively. A set of strategies $(\boldsymbol{l}_n^*, \boldsymbol{p}^{\mathrm{B}^*})$ constitutes the Stackelberg equilibrium of Γ if it satisfies the following inequalities:

$$U_n(\boldsymbol{l}_n^*, \boldsymbol{l}_{-n}^*) \le U_n(\boldsymbol{l}_n, \boldsymbol{l}_{-n}^*) \tag{29}$$

and

$$U(\boldsymbol{p}^{\mathrm{B}^*}) \le U(\boldsymbol{p}^{\mathrm{B}}). \tag{30}$$

Thus, once all residential user demands are at the Stackelberg equilibrium in Γ, no residential user can further minimize its cost function by deviating to other strategies. The problem is formulated to find the set of strategies $(\boldsymbol{l}_n^*, \boldsymbol{p}^{\mathrm{B}^*})$.

To develop algorithms for distributed community energy management, we first consider the use of weather-based LSTM for forecasting the renewable energy generation. Demand information is exchanged and appliance scheduling is optimized to reach the Stackelberg equilibrium or near Stackelberg equilibrium by several game rounds. In each round, the CA applies Q-learning to solve (27). The residential users in the community apply multiagent Q-learning to schedule their appliances and battery to solve (18). After reaching the Stackelberg equilibrium, they execute their individual scheduling in the current time slot, transition to the next time slot, and reform a new Stackelberg game. Details of the proposed method for community energy management are given in the following subsections.

A. Weather-Based LSTM for Renewable Energy Generation

To deal with the uncertainty of renewable energy, an LSTM based method using weather forecast information is employed to predict the future renewable energy generation. LSTM networks are well-suited to making predictions for temporal data because they use a memory cell to remember previous important states and learn to reset the cell for unimportant features.

In each time slot, the input vector of the trained LSTM is the past renewable energy generation data and weather forecast information w, including the sun hour, cloud cover and humidity. The output is the future renewable energy generation. This predicted information will be used for scheduling home appliances. The prediction can be symbolically expressed as

$$\begin{bmatrix} p_{t+1}^{\mathrm{PV}} & p_{t+2}^{\mathrm{PV}} & \cdots & p_{t+T}^{\mathrm{PV}} \end{bmatrix} = LSTM(\begin{bmatrix} p_{t-T+1}^{\mathrm{PV}} \cdots & p_t^{\mathrm{PV}} & w \end{bmatrix}). \tag{31}$$

B. Multiagent Q-Learning

After obtaining the forecast PV generation, we exploit RL to find the best appliance scheduling. RL deals with how an agent chooses a proper action to minimize a cumulative cost (or maximize a cumulative reward) in an uncertain environment. RL includes three major components: states, actions, and costs. States are the representation of the status of the agent in the environment. Actions are what the RL agent can act to the environment. A cost is the feedback the agent receives from the environment for the action taken.

Let $\mathcal{S}$ denote the state set and $\mathcal{A}$ denote the action set. The state of time slot t is s_t where $s_t \in \mathcal{S}$. Given s_t, the RL agent selects an action a_t from action set $\mathcal{A}$. After the agent takes an action a_t, the environment will feedback the cost r_t and transition to next state s_{t+1}. A policy $\pi(\cdot)$ is a mapping from the state space to action space. The goal of an agent is to find an optimal policy π_* that minimizes the expected cumulative cost. Given a state s_t, an action a_t and a policy π, the action-value function q_π for policy π is defined as

$$q_\pi(s_t, a_t) = \mathbb{E}_\pi\left[\sum_{i=t+1}^{T} \gamma^{i-t-1} r_{i-1} | s_t, a_t\right] \quad \forall s_t \in \mathcal{S}, \forall a_t \in \mathcal{A} \tag{32}$$

where $\gamma \in (0, 1]$ is a discount factor. The optimal action-value function is denoted as $q_*(s_t, a_t)$.

Q-learning [29] is a value-based RL algorithm that can be used to seek $q_*(s_t, a_t)$. Q-learning constructs a Q-table containing Q-value $Q(S_t, A_t)$ for each state-action pair (S_t, A_t). Q-value $Q(S_t, A_t)$ estimates the expected cumulative reward of action A_t at state S_t. The cost function $R(S_t, A_t)$ is the feedback from the environment. Q-value $Q(S_t, A_t)$ approximating $q_*(s_t, a_t)$ can be updated by

$$\begin{aligned} Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1}(S_t, A_t) \\ + \gamma \min_a Q(S_{t+1}, a) - Q(S_t, A_t)]. \end{aligned} \tag{33}$$

In Q-learning, ε-greedy action selection is generally used, i.e., greedy action

$$A_t = \arg\min_a Q(S_t, a) \tag{34}$$

is selected with probability $1 - \varepsilon$ and a random action is selected with probability ε.

At the middle layer in Fig. 1, the following setting was used for applying Q-learning: state S_t is designed as (t, B_t), action $A_t = p_t^{\mathrm{B}}$ represents the charging or discharging of the battery,

Algorithm 1: Q-Learning for CA's Objective Optimization at Time. h

Input: whole community demand profile $\boldsymbol{d}_h$, forecast PV generation profile $\boldsymbol{p}_h^{\text{PV}}$.
Output: the community storage dispatch profile $\boldsymbol{p}_h^{\text{B}}$ (greedy action selection with respect to Q^{CA}).

Create a planning model with $\boldsymbol{p}_h^{\text{PV}}$ and $\boldsymbol{d}_h$.
for episode **do**
  Initialize state S_h.
  for $t = 0 : T - h$ **do**
    Choose action A_{h+t} (i.e., p_{h+t}^{B}) for the current state S_{h+t} by ε-greedy action selection.
    Take action A_{h+t}, observe reward R_{h+t+1}^{ag} and next state S_{h+t+1}.
    Update $Q^{CA}(S_{h+t}, A_{h+t})$ using (35).

and cost R_{t+1}^{CA} is the electricity cost. To minimize the total cost of the community, the update rule is used:

$$Q^{CA}(S_t, A_t) \leftarrow Q^{CA}(S_t, A_t) + \alpha[R_{t+1}^{\text{CA}}(S_t, A_t) + \gamma \min_a Q^{CA}(S_{t+1}, a) - Q^{CA}(S_t, A_t)]. \tag{35}$$

Algorithm 1 presents the control algorithm for scheduling the CA's battery $\boldsymbol{p}_h^{\text{B}}$. The CA gathers demand load profile of all residential users in the coming time slot $\boldsymbol{d}_h := [\, d_h \; d_{h+1} \; \cdots \; d_T \,]$, where $d_h = \sum_{n \in \mathcal{C}} l_{n,h}$, and forecasts PV generation $\boldsymbol{p}_h^{\text{PV}} := [\, p_h^{\text{PV}} \; p_{h+1}^{\text{PV}} \; \cdots \; p_T^{\text{PV}} \,]$ using (31). After that, the CA creates a planning model with $\boldsymbol{d}_h$ and $\boldsymbol{p}_h^{\text{PV}}$. The storage system agent will learn episode by episode. Finally, the agent outputs the greedy action selection $\boldsymbol{p}_h^{\text{B}} := [\, p_h^{\text{B}} \; p_{h+1}^{\text{B}} \; \cdots \; p_T^{\text{B}} \,]$. The Q-table is then saved and used in the next time slot.

For residential users, let $\boldsymbol{P}_{n,h}$ denote the power control vector of all appliances and battery in residence n from time slot h to final time slot T. The power control vector $\boldsymbol{P}_{n,h}$ is defined as

$$\boldsymbol{P}_{n,h} = \begin{bmatrix} \boldsymbol{u}_{n,1,h}^{\text{shift}} \\ \vdots \\ \boldsymbol{u}_{n,I_n,h}^{\text{shift}} \\ \boldsymbol{p}_{n,1,h}^{\text{con}} \\ \vdots \\ \boldsymbol{p}_{n,J_n,h}^{\text{con}} \\ \boldsymbol{p}_{n,h}^{\text{EV}} \\ \boldsymbol{p}_{n,h}^{\text{B}} \end{bmatrix} = \begin{bmatrix} u_{n,1,h}^{\text{shift}} & u_{n,1,h+1}^{\text{shift}} & \cdots & u_{n,1,T}^{\text{shift}} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n,I_n,h}^{\text{shift}} & u_{n,I_n,h+1}^{\text{shift}} & \cdots & u_{n,I_n,T}^{\text{shift}} \\ p_{n,1,h}^{\text{con}} & p_{n,1,h+1}^{\text{con}} & \cdots & p_{n,1,T}^{\text{con}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,J_n,h}^{\text{con}} & p_{n,J_n,h+1}^{\text{con}} & \cdots & p_{n,J_n,T}^{\text{con}} \\ p_{n,h}^{\text{EV}} & p_{n,h+1}^{\text{EV}} & \cdots & p_{n,T}^{\text{EV}} \\ p_{n,h}^{\text{B}} & p_{n,h+1}^{\text{B}} & \cdots & p_{n,T}^{\text{B}} \end{bmatrix}. \tag{36}$$

At the demand side in Fig. 1, agents for controlling batteries and all appliances cooperatively work in a sequential way. Each agent has its own Q-table that records the cumulative cost of the appliance, leading to multiagent Q-learning. The update rule is

$$Q_{n,k}(S_t, A_t) \leftarrow Q_{n,k}(S_t, A_t) + \alpha[R_{n,k,t+1}(S_t, A_t) + \gamma \min_a Q_{n,k}(S_{t+1}, a) - Q_{n,k}(S_t, A_t)] \tag{37}$$

Algorithm 2: Multiagent Q-Learning for Residential User n at Time. h

Input: demand of other residential users $\boldsymbol{l}_{-n,h}$, forecast PV generation profile $\boldsymbol{p}_{n,h}^{\text{PV}}$, $D_{n,t}^{\text{non}}$, $T_{n,i,\text{start}}^{\text{shift}}$, $T_{n,i,\text{end}}^{\text{shift}}$, $w_{n,i}^{\text{shift}}$, $D_{n,i,t}^{\text{shift}}$, $T_{n,j,\text{start}}^{\text{con}}$, $T_{n,j,\text{end}}^{\text{con}}$, $D_{n,j,\text{min}}^{\text{con}}$, $D_{n,j,\text{max}}^{\text{con}}$, $T_{n,\text{start}}^{\text{EV}}$, $T_{n,\text{end}}^{\text{EV}}$, $D_{n,\text{min}}^{\text{EV}}$, $D_{n,\text{max}}^{\text{EV}}$, B_n^{min}, B_n^{max}, dissatisfaction coefficients $\beta_{n,i}^{\text{shift}}$, $\beta_{n,j}^{\text{con}}$, β_n^{EV}.
Output: residential user's power demand $\boldsymbol{l}_{n,h}$ with $\boldsymbol{P}_{n,h}$ defined in (13).

Sort the appliances by dissatisfaction coefficient $\beta_{n,i}^{\text{shift}}$, $\beta_{n,j}^{\text{con}}$, β_n^{EV} in a descending order.
Create a planning model with $\boldsymbol{l}_{-n,h}$ and available PV generation $\boldsymbol{p}_{n,h}^{\text{PV}}$.
for each appliance agent k **do**
  for episode **do**
    Initialize state S_t.
    for $t = 0 : T - h$ **do**
      Choose action A_{h+t} for the current state S_{h+t} by ε-greedy action selection.
      Take action A_{h+t}, observe the current cost $R_{n,k,h+t+1}$ and the next state S_{h+t+1}.
      Update the Q-value $Q_{n,k}(S_{h+t}, A_{h+t})$ using (37).
  Obtain the scheduling of the agent from the greedy action selection with respect to $Q_{n,k}$, i.e., $\boldsymbol{u}_{n,i,h}^{\text{shift}}$, $\boldsymbol{p}_{n,h}^{\text{B}}$.
  Update available PV generation $\boldsymbol{p}_{n,h}^{\text{PV}}$ by performing the scheduling of appliance agent k

where $Q_{n,k}$ represents the Q-value of appliance k, which represents a shiftable appliance, controllable appliance, EV or ESS.

Table I lists our designs of states, actions and costs of different agents controlling the appliances and battery. The state of a shiftable appliance provides information about the current time slot and left working time. Two actions "on" and "off" can be selected, where action "on" is associated with the electricity cost and action "off" is associated with user dissatisfaction. The state for the controllable appliances and EV is the current time slot. The actions are their flexible power rating. Their cost functions are the electricity spending and user dissatisfaction. The state, action and cost function of the battery are the same as those of the CA.

Algorithm 2 is a multiagent Q-learning algorithm for all appliances and the battery in residence. Dissatisfaction coefficients $\beta_{n,i}^{\text{shift}}$, $\beta_{n,j}^{\text{con}}$, and β_n^{EV} are sorted in a descending order. A planning model is composed by the demand of other residential users $\boldsymbol{l}_{-n,h} := [\, l_{-n,h} \; l_{-n,h+1} \; \cdots \; l_{-n,T} \,]$ and forecast PV generation $\boldsymbol{p}_{n,h}^{\text{PV}} := [\, p_{n,h}^{\text{PV}} \; p_{n,h+1}^{\text{PV}} \; \cdots \; p_{n,T}^{\text{PV}} \,]$. The battery agent learns through the planning model and obtains the best scheduling. The battery agent executes and updates the available PV generation $\boldsymbol{p}_{n,h}^{\text{PV}}$ using the planning model, which helps the other appliance agents to learn. In the end, the residential user sends its own demand load $\boldsymbol{l}_{n,h} := [\, l_{n,h} \; l_{n,h+1} \; \cdots \; l_{n,T} \,]$ to the CA with greedy selection action vector $\boldsymbol{P}_{n,h}$.

TABLE I
DESIGNS OF STATE, ACTION AND COST FOR EACH AGENT

Agnet	S_t	A_t	R_{t+1}
Shiftable Appliance i	$(t,d^{\text{shift}}_{n,i,t})$	$u^{\text{shift}}_{n,i,t} = \begin{cases} 0 \text{ represents "off"} \\ 1 \text{ represents "on"} \end{cases}$	$\lambda_t u^{\text{shift}}_{n,i,t} D^{\text{shift}}_{n,i,t} \Delta t + g^{\text{shift}}_{n,i,t}$
Controllable Appliance j	t	power rating $p^{\text{con}}_{n,j,t}$	$\lambda_t p^{\text{con}}_{n,j,t} \Delta t + g^{\text{con}}_{n,i,t}$
EV	t	power charging $p^{\text{ev}}_{n,t}$	$\lambda_t p^{\text{EV}}_{n,t} \Delta t + g^{\text{EV}}_{n,t}$
Battery	$(t,B_{n,t})$	$p^{\text{B}}_{n,t} = \begin{cases} \text{charging}, & p^{\text{B}}_{n,t} \geq 0 \\ \text{discharging}, & p^{\text{B}}_{n,t} < 0 \end{cases}$	$\lambda_t p^{\text{B}}_{n,t} \Delta t$
CA Battery	(t,B^{ag}_t)	$p^{\text{B}}_t = \begin{cases} \text{charging}, & p^{\text{B}}_t \geq 0 \\ \text{discharging}, & p^{\text{B}}_t < 0 \end{cases}$	$\lambda_t p^{\text{CA}}_t \Delta t$

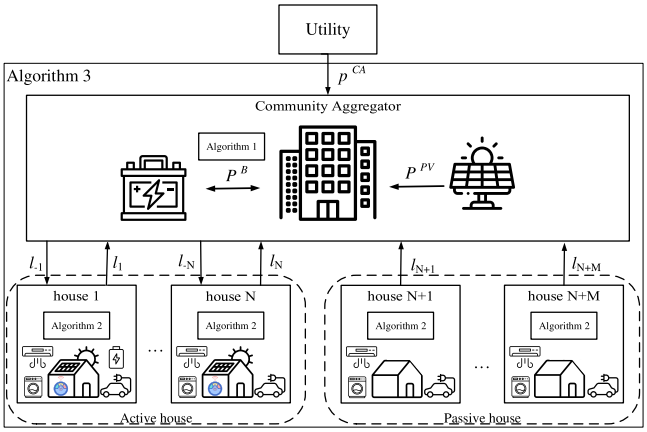

Fig. 2. Roles of the proposed algorithms in community energy management.

Algorithm 3 presents the proposed distributed community energy management algorithm. In each time slot h, the CA and the residential users reach the Stackelberg equilibrium through several game rounds. At each game round r, residential user n forecasts its own PV generation using (31), determines its best appliances' scheduling to minimize its cost function (18) with demand of other residential users $\boldsymbol{l}_{-n,h}{}^{(r-1)}$, and produces updated demand load $\boldsymbol{l}_{n,h}{}^{(r)}$ and notifies the CA of that information. The CA receives the demand of all residential users, forecasts its PV generation using (31), schedules its battery storage dispatch $\boldsymbol{p}^{\text{B}}_h{}^{(r)}$ using Algorithm 1, and broadcasts the new $\boldsymbol{l}_{-n,h}{}^{(r)}$ to all residential users. This interaction continues until the Stackelberg equilibrium is reached, i.e., the cost function of the CA does not change. A new Stackelberg game between the CA and residential users will form in the next time slot.

Fig. 2 visualizes the relationship between the proposed algorithms. Algorithm 1 controls the ESS of the CA on the basis of the community demand profile and forecast PV generation at the CA. Algorithm 2 determines the power demands and power control profiles for all appliances. Algorithm 3 iteratively and dynamically adjusts the decisions made by Algorithms 1 and 2 so that a near Stackelberg equilibrium can be reached. From the perspective of operating regions, Algorithm 1 is implemented at the CA, Algorithm 2 is implemented in each residence, and Algorithm 3 involves the interactions between the CA and all residences.

IV. NUMERICAL ANALYSIS

This section examined the effectiveness of the proposed community energy management method. The scenario involved a CA, $N = 10$ active residential users, and $M = 5$ passive residential users. The CA was equipped with a 30kWp solar PV system and 62.5 kWh battery system. Suppose that five active residences were equipped with an ESS to balance the load in response to price variance and solar generation; the other five active residences were equipped with solar generation. The active residences were installed a solar PV system and ESS. See Table II for the setting. The charging/discharging efficiency coefficients $\eta_n = 0.9$ in (10) and $\eta^{\text{ag}} = 1.1$ in (19) were set. The degradation costs of battery C^{deg}_n and C^{deg} were set to 0.2. The overprice $\tau = 1000$ was applied to force the CA to meet the constraints. The weight $w_n = 0.5$ in (16) was used. Each residence was equipped with two non-shiftable appliances (refrigerator and cooker), two shiftable appliances (washing machine and dishwasher), two controllable appliances (water heater and air conditioner), and an EV [21]; Table III shows the relevant parameters. In (25), λ_t was set as following [17], [21]:

$$\lambda_t = \begin{cases} 0.6 + 0.045 p^{\text{CA}}_t & t = 1, 2, \ldots, 8 \\ 0.8 + 0.06 p^{\text{CA}}_t & t = 9, 10, \ldots, 24 \end{cases}$$

For the parameters in Q-learning, discount rate $\gamma = 0.99$ and learning rate $\alpha = 0.1$ in (33), and $\varepsilon = 0.1$ were set.

Real-world data were applied for training our weather-based LSTM method that used weather information such as sun hour, cloud cover, and humidity. The hourly solar generation was collected form PJM [30]. The weather forecasting information was gathered from World Weather Online [31]. Fig. 3 presents a sample of 24-hour power demand of a residential user starting from time slot labeled as 07:00 (time period from 07:00 to 07:59) to the time slot labeled as 06:00 (next-day time period from 06:00 to 06:59). Peak demand was approximately 8 kW and occurred in time slot 13:00. Renewable generation began in time slot 9:00 and ended in time slot 19:00, with peak generation during time slots 14:00 and 15:00. When a bar representing "battery" meets the demand curve, a charging event occurs. For example, the battery was charged in time slots 9:00, 13:00–16:00, and 20:00. The charging activities during 13:00–16:00 coincided with the peak renewable generation. When a bar representing "battery" does not meet the demand curve, a discharging event occurs. As shown in the figure, the battery was discharged in time slots 07:00, 18:00, 19:00, 23:00, 00:00, 03:00, 04:00, and 06:00.

Algorithm 3: Proposed Multiagent Method for Community Energy Management.

Input: $D^{\text{non}}_{n,t}$, $T^{\text{shift}}_{n,i,\text{start}}$, $T^{\text{shift}}_{n,i,\text{end}}$, $w^{\text{shift}}_{n,i}$, $D^{\text{shift}}_{n,i,t}$, $T^{\text{con}}_{n,j,\text{start}}$, $T^{\text{con}}_{n,j,\text{end}}$, $D^{\text{con}}_{n,j,\min}$, $D^{\text{con}}_{n,j,\max}$, $T^{\text{EV}}_{n,\text{start}}$, $T^{\text{EV}}_{n,\text{end}}$, $D^{\text{EV}}_{n,\min}$, $D^{\text{EV}}_{n,\max}$, $B^{\min}_{n}$, $B^{\max}_{n}$, dissatisfaction coefficients $\beta^{\text{shift}}_{n,i}, \beta^{\text{con}}_{n,j}, \beta^{\text{EV}}_{n}$ of each residential user n, $B^{\min}$, $B^{\max}$ of CA and electricity pricing coefficient e_t, f_t.

Output: energy scheduling of home appliances for each residential user $\boldsymbol{P}_n$ and CA storage dispatch $\boldsymbol{p}^{\text{B}}$.

Initialize aggregator Q-table $Q^{CA}(s,a)$ and $Q_{n,k}(s,a)$ of all appliances in each residence arbitrarily.
For $h = 1, 2, \ldots T$ **do**
  Initialize $U^{(0)}$ and round $r = 1$.
  CA receives demand load $\boldsymbol{l}^{(0)}_{n,h}$ from all residential users and calculates initial total demand profile $\boldsymbol{d}^{(0)}_{h}$.
  CA forecasts PV generation profile $\boldsymbol{p}^{\text{PV}}_{h}$ produced by (31).
  CA schedules the community storage dispatch profile $\boldsymbol{p}^{\text{B}}_{h}$ produced by **Algorithm 1**, and calculates community load profile $\boldsymbol{p}^{(r)}_{h}$ to minimize $U^{(r)}$.
  while $||U^{(r)} - U^{(r-1)}|| > \xi$ **do**
    $r \leftarrow r + 1$
    CA broadcasts the community load profile $\boldsymbol{p}^{(r-1)}_{h}$ to all residential users.
    for residential user $n \in \mathcal{C}$ **do**
      Residential user receives $\boldsymbol{l}^{(r-1)}_{-n,h}$ $(\boldsymbol{p}^{(r-1)}_{h} - \boldsymbol{l}_{n,h}{}^{(r-1)})$ broadcasted by CA.
      Residential user forecasts its PV generation profile $\boldsymbol{p}^{\text{PV}}_{n,h}$ produced by (31) and produces the scheduling for all appliances using **Algorithm 2.**
      Residential user sends its demand load profile $\boldsymbol{l}^{(r)}_{n,h}$ to CA.
    CA receives all demand load and calculates community total demand profile $\boldsymbol{d}^{(r)}_{h}$.
    CA forecasts PV generation profile $\boldsymbol{p}^{\text{PV}}_{h}$ produced by (31).
    CA schedules the community storage dispatch profile $\boldsymbol{p}^{\text{B}}_{h}$ to minimize $U^{(r)}$ using **Algorithm 1** and calculates community load profile $\boldsymbol{p}^{(r)}_{h}$.
  CA executes the community storage dispatch $\boldsymbol{p}^{\text{B}}_{h}$ and residential users perform energy scheduling of all appliances.

TABLE II
SOLAR GENERATION AND BATTERY OF RESIDENCE

Active Residence Profile				
Residence	Solar PV (kWp)	Battery (kW)	Residence	Solar PV (kWp)
Residence 1, 2	4	9	Residence 6, 7	4
Residence 3, 4, 5	6	13.4	Residence 8, 9, 10	6

TABLE III
PARAMETERS FOR HOUSE APPLIANCES

Appliance ID	Dissatisfaction coefficient	Power rating (kWh)	Using time
REFG	-	0.5	24 h
CK	-	1.2	rand(7, 9), rand(12, 14), rand(18, 20)
WM	1.4	0.7	rand(9, 11) to rand(17, 21), duration=rand(3, 4)
DW	1.0	0.3	rand(12, 14) to rand(23, 25), duration=rand(2, 3)
WH	3.0	1.5	rand(7, 9) to rand(29, 30)
AC	2.0	1.4	rand(7, 9) to rand(29, 30)
EV	1.5	1.44	rand(19, 22) to rand(29, 30)

REFG: refrigerator; CK: cooker; WH: washing machine; DW: dishwasher; WH: water heater; AC: air conditioner; EV: electric vehicle.

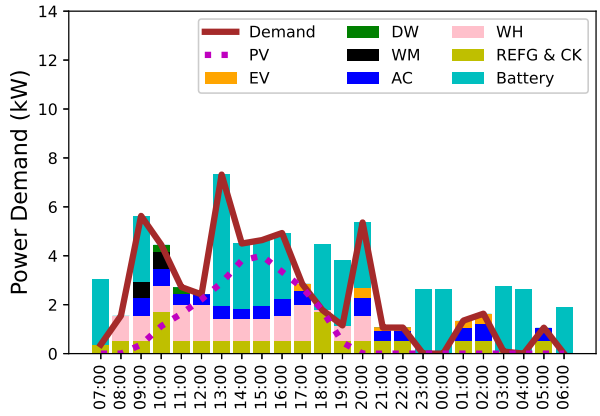

Fig. 3. Residential user demand response resulting from the proposed multiagent method for community energy management.

A. Comparison With Learning Based Methods for Community Energy Management

The proposed method for community energy management was able to mitigate peak rebounds. Figs. 4(a) and (b) present the community demand resulting from the proposed method for community energy management and the energy management by individual HEMSs (a single-agent RL based method), respectively. For the individual HEMS, high load occurred in time slots 08:00 and 09:00, and peak load was 56.47 kW. This may be due the predicted low prices by an HEMS in that period. The HEMS thus shifted shiftable appliances and turned up the power of controllable appliances accordingly. This practice caused a peak rebound and much burden on the gird, yielding a high load. By contrast, coordinated behaviors encouraged by the proposed method had a peak load of 39.07 kW, which was 30.8% peak reduction as compared with the energy management by individual HEMSs.

Table IV shows the average residential user costs in four seasons. The proposed method outperformed the single-agent method in all cases. The single-agent method can synchronously increase the demand. Such uncoordinated behaviors can increase the market price and thus incur a larger average user cost than coordinated behaviors encouraged by our multiagent method.

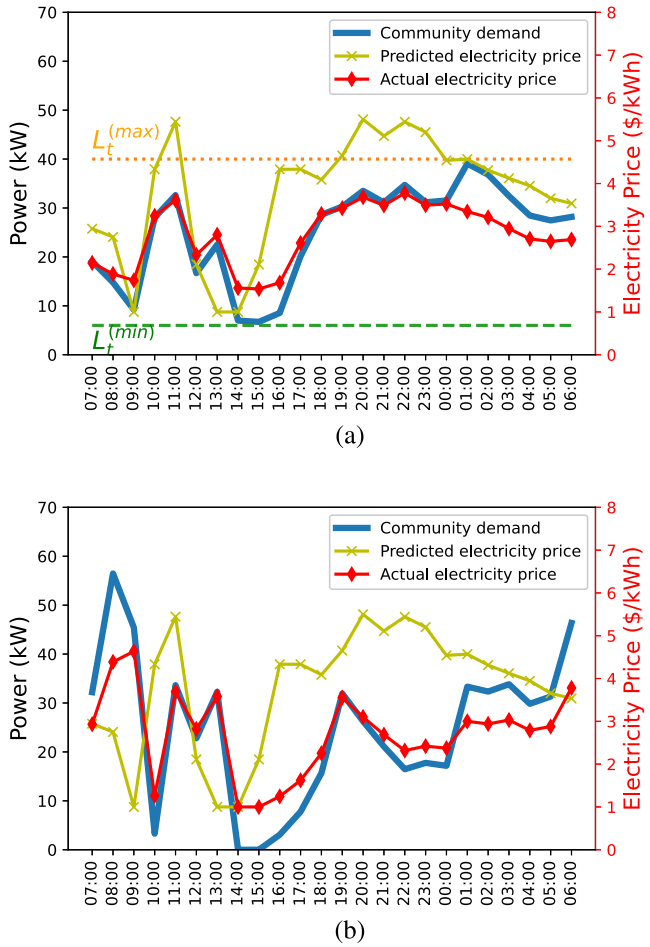

Fig. 4. Community demand response resulting from (a) the proposed community energy management and (b) individual HEMSs. Energy management by individual HEMSs (single-agent RL based method) led high demand in time slots 08:00 and 09:00 because a low price period was expected, yielding a peak rebound. The proposed method for community energy management mitigated such a rebound.

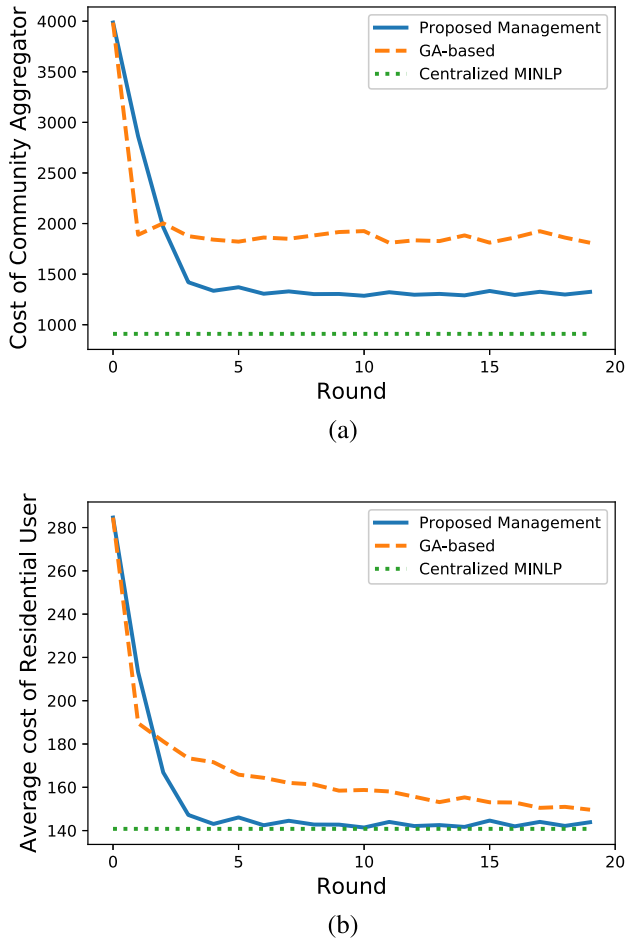

Fig. 5. Convergence of comparable algorithms to the Stackelberg equilibrium given perfect information about renewable generation at (a) CA and (b) residential users. Centralized MINLP provided an ideal level of performance but required information about all residential users to be processed at the CA, raising a privacy concern. The proposed community energy management method achieved a lower aggregator cost and average residential user cost than GA-based optimization within 20 game rounds (the number of iterations used to solve the Stackelberg game); it approximately attained the ideal average residential user cost upon increasing the game rounds.

TABLE IV
PERFORMANCE COMPARISON OF MULTIAGENT AND SINGLE-AGENT METHODS FOR COMMUNITY ENERGY MANAGEMENT

Performance metrics \ Methods	Single-agent optimization	Proposed community energy management
Average residential user cost on Spring day (dollars)	217.55	202.64 (-6.85%)
Average residential user cost on Summer day (dollars)	206.63	172.82 (-16.36%)
Average residential user cost on Fall day (dollars)	192.06	173.96 (-9.42%)
Average residential user cost on Winter day (dollars)	212.66	199.74 (-6.1%)

The average cost reduction for all seasons by the proposed method was 9.68%.

B. Comparison With Optimization Based Methods for Community Energy Management

The proposed method for community energy management was further compared with the centralized MINLP [19], [32] and genetic algorithm (GA) based solution method [21]. In Fig. 5, we assume future renewable generation is available and can be used in energy scheduling optimization. As such, the centralized MINLP produced an ideal level of performance while requiring information about all residential users to be processed at the CA and hence, raising a privacy concern. The proposed and GA based solution methods converged after a few game rounds (the number of iterations used to solve the Stackelberg game). The proposed multiagent method for energy management was better than the GA based solution method; it also approximately attained the ideal average residential user cost upon increasing the game rounds.

In practice, however, future renewable generation can only be estimated, degrading the performance of any optimization methods such as the centralized MINLP. The weather-based LSTM method was applied to estimate the renewable generation and resulted in prediction errors. Table V presents the average

TABLE V
PERFORMANCE COMPARISON WITH RENEWABLE UNCERTAINTY

Performance metrics \ Methods	Given Predicted Renewable Energy Generation		
	Centralized MINLP [19]	GA-based optimization [21]	Proposed community energy management
CA cost	1373.4 (-16.5%)	2524.1 (53.4%)	1645.9
Average residential user cost	168.3 (2.7%)	178.4 (8.9%)	163.8
Standard deviation of residential user cost	614.7 (56%)	642.3 (63.1%)	393.8
Maximum aggregator load (kW)	41.5 (1.72%)	75.6 (85.3%)	40.8

costs, standard deviations, and peak load at the CA. The standard deviation can be used as a robustness measure; a smaller standard deviation means that a method can produce a more consistent result. The peak load should be as small as possible for a stable system and less expensive capacity to be constructed. The centralized MINLP had the lowest CA cost, 16.5% cost reduction from the proposed energy management. This was consistent with the result presented in Fig. 5 that centralized MINLP was more advantageous than our learning based method in terms of the CA's cost. However, this advantage may come from the fact that centralized MINLP considered all local information in residences as global information, which is not practical. Except for the CA cost, the proposed community energy management method outperformed centralized MINLP in terms of the average residential cost by 2.7%, standard deviation by 56%, and the peak load at the CA by 1.72%.

V. CONCLUSION

Research on community energy management is increasing because of the ubiquity of EVs and renewable energy as well as the growth of the smart home industry. Moreover, MARL has been applied to energy management because of its ability to address uncertainty resulting from the uncoordinated behaviors of energy users. This study therefore focused on MARL for energy management in a residential community and investigated two critical aspects of community energy scheduling: peak rebounds and the uncertainty of renewable energy generation. Peak rebounds place extra pressure on the grid, and the uncertainty of renewable energy generation affects the efficiency of energy scheduling. To address these problems, we propose a community energy management method, in which appliance scheduling is formulated as a game solved by a multiagent approach.

Evaluations in the form of numerical analyses were conducted. In these evaluations, the proposed method could address peak rebounds and the uncertainty of renewable energy generation while minimizing energy costs. By using the proposed MARL-based method, we achieved a peak load reduction of 30.8% and average cost reduction of 9.68% compared with a single-agent RL-based method. Compared with optimization methods for community energy management, the proposed method outperformed centralized MINLP in terms of the average residential cost by 2.7%, standard deviation of the cost by 56%, and the peak load of the CA by 1.72%.

The learning algorithms considered in this study used tabular solution methods, and Q-tables were learned for decision-making. One limitation of tabular solution methods is that the state space and action space must be finite and small. Our future research will investigate the use of deep RL for a community EMS that generalizes well in large state and action spaces.

REFERENCES

[1] M. Muratori and G. Rizzoni, "Residential demand response: Dynamic energy management and time-varying electricity pricing," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1108–1117, Mar. 2015.
[2] P. Siano, "Demand response and smart grids–A survey," *Renewable Sustain. Energy Rev.*, vol. 30, pp. 461–478, 2014.
[3] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Appl. Energy*, vol. 235, pp. 1072–1089, 2019.
[4] Z. Zhao, W. C. Lee, Y. Shin, and K.-B. Song, "An optimal power scheduling method for demand response in home energy management system," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1391–1400, Sep. 2013.
[5] A. Anvari-Moghaddam, H. Monsef, and A. Rahimi-Kian, "Cost-effective and comfort-aware residential energy management under different pricing schemes and weather conditions," *Energy Buildings*, vol. 86, pp. 782–793, Jan. 2015.
[6] C. Chen, J. Wang, and S. Kishore, "A distributed direct load control approach for large-scale residential demand response," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2219–2228, Sep. 2014.
[7] Y. Li, B. L. Ng, M. Trayer, and L. Liu, "Automated residential demand response: Algorithmic implications of pricing models," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1712–1721, Dec. 2012.
[8] A. Safdarian, M. Fotuhi-Firuzabad, and M. Lehtonen, "A distributed algorithm for managing residential demand response in smart grids," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2385–2393, Nov. 2014.
[9] B. Celik, R. Roche, S. Suryanarayanan, D. Bouquain, and A. Miraoui, "Electric energy management in residential areas through coordination of multiple smart homes," *Renewable Sustain. Energy Rev.*, vol. 80, pp. 260–275, Dec. 2017.
[10] D. T. Nguyen and L. B. Le, "Joint optimization of electric vehicle and home energy scheduling considering user comfort preference," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 188–199, Jan. 2014.
[11] M. H. K. Tushar, C. Assi, M. Maier, and M. F. Uddin, "Smart microgrids: Optimal joint scheduling for electric vehicles and home appliances," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 239–250, Jan. 2014.
[12] T. M. Hansen, R. Roche, S. Suryanarayanan, A. A. Maciejewski, and H. J. Siegel, "Heuristic optimization for an aggregator-based resource allocation in the smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1785–1794, Jul. 2015.
[13] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Secur. Privacy*, vol. 7, no. 3, pp. 75–77, Jun. 2009.
[14] G. Tsaousoglou, K. Steriotis, N. Efthymiopoulos, P. Makris, and E. Varvarigos, "Truthful, practical and privacy-aware demand response in the smart grid via a distributed and optimal mechanism," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3119–3130, Jul. 2020.
[15] P. Chavali, P. Yang, and A. Nehorai, "A distributed algorithm of appliance scheduling for home energy management system," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 282–290, Jan. 2014.
[16] A. Safdarian, M. Fotuhi-Firuzabad, and M. Lehtonen, "Optimal residential load management in smart grids: A decentralized framework," *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1836–1845, Jul. 2016.
[17] A.-H. Mohsenian-Rad, V. W. 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 Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[18] K. Wang, Z. Ouyang, R. Krishnan, L. Shu, and L. He, "A game theory-based energy management system using price elasticity for smart grids," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1607–1616, Dec. 2015.
[19] Z. Zhu, S. Lambotharan, W. H. Chin, and Z. Fan, "A game theoretic optimization framework for home demand management incorporating local energy resources," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 353–362, Apr. 2015.
[20] C. Li, X. Yu, W. Yu, G. Chen, and J. Wang, "Efficient computation for sparse load shifting in demand side management," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 250–261, Jan. 2017.
[21] B. Rajasekhar, N. Pindoriya, W. Tushar, and C. Yuen, "Collaborative energy management for a residential community: A non-cooperative and evolutionary approach," *IEEE Trans. Emerg. Top. Comput.*, vol. 3, no. 3, pp. 177–192, Jun. 2019.
[22] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.
[23] T. Remani, E. Jasmin, and T. I. Ahamed, "Residential load scheduling with renewable generation in the smart grid: A reinforcement learning approach," *IEEE Syst. J.*, vol. 13, no. 3, pp. 3283–3294, Sep. 2018.
[24] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.
[25] J. Yan *et al.*, "Frequency-domain decomposition and deep learning based solar PV power ultra-short-term forecasting model," *IEEE Trans. Ind. Appl.*, vol. 57, no. 4, pp. 3282–3295, Jul./Aug. 2021.
[26] S. Talpur, T. T. Lie, and R. Zamora, "Application of demand response and smart battery electric vehicles charging for capacity utilization of the distribution transformer," in *Proc. IEEE PES Innov. Smart Grid Technol. Eur.*, The Hague, South Holland, The Netherlands, 2020, pp. 479–483.
[27] M. Yu and S. H. Hong, "Incentive-based demand response considering hierarchical electricity market: A stackelberg game approach," *Appl. Energy*, vol. 203, pp. 267–279, Oct. 2017.
[28] I. Atzeni, L. G. Ordóñez, G. Scutari, D. P. Palomar, and J. R. Fonollosa, "Noncooperative day-ahead bidding strategies for demand-side expected cost minimization with real-time adjustments: A GNEP approach," *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2397–2412, May 2014.
[29] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3/4, pp. 279–292, 1992.
[30] PJM interconnection, Accessed: Jan. 27, 2022. [Online]. Available: http://www.pjm.com
[31] World Weather Online, Accessed: Jan. 27, 2022. [Online]. Available: https://www.worldweatheronline.com
[32] L. D. R. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, "Gekko optimization suite," *Processes*, vol. 6, no. 8, pp. 1–26, Jul. 2018.

Bo-Chen Lai received the B.S. degree in power mechanical engineering and the M.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2017 and 2020, respectively. His research interests include reinforcement learning algorithms and smart grids.

Wei-Yu Chiu (Member, IEEE) received the Ph.D. degree in communications engineering from National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2010. He is currently an Associate Professor of electrical engineering with NTHU. His research interests include multiobjective optimization and reinforcement learning, and their applications to control systems, robotics, and smart energy systems. He was the recipient of the Youth Automatic Control Engineering Award bestowed by Chinese Automatic Control Society in 2016, the Outstanding Young Scholar Academic Award bestowed by Taiwan Association of Systems Science and Engineering in 2017, the Erasmus+Programme Fellowship funded by European Union (staff mobility for teaching) in 2018, and Outstanding Youth Electrical Engineer Award bestowed by Chinese Institute of Electrical Engineering in 2020. From 2015 to 2018, he had been serving as an Organizer and the Chair for the International Workshop on Integrating Communications, Control, and Computing Technologies for Smart Grid (ICT4SG). He is a Subject Editor for *IET Smart Grid*.

Yuan-Po Tsai received the B.S. degree in electrical engineering from National Ocean University, Keelung, Taiwan, in 2017, and the M.S. degree in electrical engineering from National Ilan University, Ilan, Taiwan, in 2019. He is currently working toward the Ph.D. degree with National Tsing Hua University, Hsinchu, Taiwan. His research interests include cooperative control, reinforcement learning, and multi-robot systems.