# Intelligent Home Energy Management Strategy With Internal Pricing Mechanism Based on Multiagent Artificial Intelligence-of-Things

Tianyun Xu, Tao Chen, *Member, IEEE*, Ciwei Gao, *Senior Member, IEEE*, and Hongxun Hui, *Member, IEEE*

***Abstract*—Currently, an increasing number of residential customers have access to distributed flexible resources, including dispatchable or nondispatchable distributed generation and various flexible loads. Meanwhile, these distributed resources are also available to some local energy transactions based on internal pricing mechanism in energy communities. However, the local energy trading of distributed resources requires considerable computational capacity and professional knowledge for end-users, which makes it difficult to guarantee the trading willingness. Therefore, this article proposes an intelligent home energy management strategy for residential customers based on deep reinforcement learning techniques with consideration of internal pricing mechanism. Technically, the end-user sequential decision-making process in energy management and trading can be modeled as a Markov decision process using encapsulated resource physical status information and pricing preference information. In particular, this article considers the interactive relationship between different customers or decision-makers, capturing features of group intelligent decision that evolves as energy status change and internal price signal change. The simulation and demonstration of such an intelligent home energy management problem are provided with multiagent setup based on a quite new concept of artificial intelligence-of-things that could showcase the software and hardware implementation features at the same time. By sufficient cosimulation experimental studies, this article found that residential customers can achieve a significant improvement in their economic benefit and decision-making efficiency.**

***Index Terms*—Artificial intelligence-of-Things (AIoT), deep reinforcement learning (DRL), electricity pricing mechanism, home energy management.**

## NOMENCLATURE

*Variables and Parameters of Device Models*

| | |
|---|---|
| $i$ | Index of prosumers. |
| $t$ | Temporal index. |
| $\Delta t$ | Time interval. |
| $T^{\text{in}}_{i,t}/T^{\text{out}}_{i,t}$ | Indoor and outdoor temperature of prosumer $i$ at step $t$ (°C). |
| $\overline{T^{\text{in}}}/\underline{T^{\text{in}}}$ | Maximum and minimum comfort levels of the indoor temperature (°C). |
| $R_{\text{in}}$ | Thermal resistance of indoor rooms (°C/kW). |
| $C_{\text{in}}$ | Heat capacity of indoor air (kJ/°C). |
| $P^{\text{ac}}_{i,t}$ | Electrical power of AC of prosumer $i$ at step $t$ (kW). |
| $\eta_{\text{ac}}$ | Efficiency of AC. |
| $\overline{P_{i,\text{ac}}}$ | Rated power of AC of prosumer $i$ (kW). |
| $A^{\text{ac}}_{i,t}$ | Value variable indicates the cooling($-1$), heating(1), or shutdown(0) of AC of prosumer $i$ at step $t$. |
| $T^{\text{water}}_{i,t}$ | Temperature of the hot water in EWH of prosumer $i$ at step $t$ (kW). |
| $\underline{T^{\text{water}}}$ | Minimum comfort levels of the water (°C). |
| $\overline{T^{\text{water}}}$ | Maximum comfort levels of the water (°C). |
| $T^{\text{cold}}$ | Cold water temperature to EWH (°C). |
| $V_{\text{cistern}}$ | Water heater volume(L). |
| $V^{\text{hot}}_{i,t}$ | Amount of hot water consumed by EWH of prosumer $i$ at step $t$ (L). |
| $C_e$ | Energy conversion factor (kJ/kWh). |
| $\eta_{\text{water}}$ | Efficiency of EWH. |
| $C^{\text{water}}$ | Specific heat capacity of water (kJ/kg.°C). |
| $P^{\text{water}}_{i,t}$ | Electrical power of EWH of prosumer $i$ at step $t$ (kW). |
| $\overline{P_{i,\text{water}}}$ | Rated power of EWH of prosumer $i$ (kW). |
| $A^{\text{water}}_{i,t}$ | Binary variables indicate whether the EWH of prosumer $i$ is heated(1) or not(0) at step $t$. |
| $\text{SOC}_{i,t}$ | Battery charged state of EV of prosumer $i$ at step $t$. |
| $P^{\text{ch}}_{i,t}/P^{\text{dis}}_{i,t}$ | Charge and discharge power of EV of prosumer $i$ at step $t$ (kW). |
| $\eta^{\text{ch}}/\eta^{\text{dis}}$ | Charging and discharging efficiency of EV. |
| $C_{\text{bat}}$ | Total battery capacity of EV (kWh). |
| $\text{SOC}^{\text{tr}}_{i,t}$ | Commuting energy consumption of EV of prosumer $i$ at step $t$. |
| $\overline{P_{i,\text{ev}}}$ | Maximum charging and discharging power of EV of prosumer $i$ (kW). |
| $A^{\text{ev}}_{i,t}$ | Binary variables indicate whether the EV of prosumer $i$ can charge and discharge(1) or not(0) at step $t$. |
| $\text{SOC}_{\text{min}}$ | Minimum state of charge of EV. |
| $\text{SOC}_{\text{max}}$ | Maximum state of charge of EV. |

Manuscript received 25 November 2022; revised 21 May 2023 and 24 August 2023; accepted 9 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 52107079, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20210243, and in part by the Open Research Project Program of the State Key Laboratory of Internet of Things for Smart City (University of Macau) under Grant SKL-IoTSC(UM)-2021-2023/ORPF/A14/2022. *(Corresponding author: Tao Chen.)*

Tianyun Xu, Tao Chen, and Ciwei Gao are with the School of Electrical Engineering, Southeast University, Nanjing 211189, China (e-mail: 220222656@seu.edu.cn; taoc@seu.edu.cn; ciwei.gao@seu.edu.cn).

Hongxun Hui is with the State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macau, Macau 211189, China (e-mail: hongxunhui@um.edu.mo).

Digital Object Identifier 10.1109/JSYST.2023.3324795

| | |
|---|---|
| $\alpha_{\text{ev}}/\beta_{\text{ev}}$ | Start time and end time of commute. |
| $\text{SOC}_\alpha$ | Power to start commute. |
| $\text{SOC}_H$ | Minimum power for commuting. |
| $P^s_{i,t}$ | Net load of prosumer $i$ at step $t$ (kW). |
| $P^{\text{fix}}_t i,t$ | Nonshiftable load of prosumer $i$ at step $t$ (kW). |
| $P^{\text{pv}}_{i,t}$ | Photovoltaic production of prosumer $i$ at step $t$ (kW). |
| $P^{\text{producer}}_t$ | Total net generation at step $t$ (kW). |
| $P^{\text{consumer}}_t$ | Total net demand at step $t$ (kW). |
| $\lambda^{\text{local,buy}}_t$ | Internal electricity price of buying. |
| $\lambda^{\text{local,sell}}_t$ | Internal electricity price of selling. |
| $\lambda^{\text{grid,buy}}_t$ | External electricity price of buying. |
| $\lambda^{\text{grid,sell}}_t$ | External electricity price of selling. |
| $T^{\text{in}}_{\text{comf}}$ | Set comfortable indoor temperature (°C). |
| $T^{\text{water}}_{\text{comf}}$ | Set comfortable water temperature (°C). |
| $N^{\text{ac}}_{\text{pen}}/N^{\text{ewh}}_{\text{pen}}$ | Times of indoor/water temperature out of comfortable range. |

*Variables and Parameters of Markov Decision Process*

| | |
|---|---|
| $S_{i,t}$ | State space. |
| $a_{i,t}$ | Action space. |
| $a^{\text{av}}_{i,t}$ | Action of AC of prosumer $i$ at step $t$. |
| $a^{\text{water}}_{i,t}$ | Action of EWH of prosumer $i$ at step $t$. |
| $a^{\text{ev}}_{i,t}$ | Action of EV of prosumer $i$ at step $t$. |
| $R^{\text{cost}}_{i,t}$ | Reward of trading electricity of prosumer $i$ at step $t$. |
| $R^{\text{ac}}_{i,t}$ | Reward of indoor temperature of prosumer $i$ at step $t$. |
| $R^{\text{water}}_{i,t}$ | Reward of water temperature of prosumer $i$ at step $t$. |
| $R^{\text{ev}}_{i,t}$ | Reward of SOC level of prosumer $i$ at step $t$. |
| $R^s_{i,t}$ | Total reward of prosumer $i$ at step $t$. |
| $\text{cost}^{\text{ac}}_{i,t}$ | Economic index of AC of prosumer $i$. |
| $\text{cost}^{\text{water}}_{i,t}$ | Economic index of EWH of prosumer $i$. |
| $\text{cost}^{\text{ev}}_{i,t}$ | Economic index of EV of prosumer $i$. |

## I. Introduction

THE modern power system with expectation of less fossil fuel resources and higher renewable energy penetration rate require new efficient energy management strategies [1] and transactive energy solutions [2], [3] to support the development of a much more flexible energy ecosystem. Many types of microgrids, distributed energy community and shared energy projects in different regions are actively exploring new paradigm of demand-side energy interactions and transactions, for example, Brooklyn Microgrid Project, Olympic Peninsula Tradable Energy Project, Sonnen Community Local Energy Trading Project in Germany, Powerpeers Shared Energy Project in the Netherlands, and many others [4]. Although the concept of distributed energy trading is progressing rapidly during the last decade the resource allocation and pricing in the electricity market are mostly still based on the traditional hierarchical top-down framework of power system management [5], which makes prosumers remain passive in local energy trading. This has led to problems, such as high transaction costs, low efficiency, and slow regulatory responses [6]. On the other hand, the pricing signal at wholesale market level is hard to broadcast at retail market level, which especially affects downstream energy end-users to guide their consumption behavior on the basis of system operation requirements [7], [8]. Therefore, some works try to conceptualize the idea of transactive energy or transactive control [9] to link the pricing mechanism to the energy status of flexible distributed resources, pushing forward subtle energy flow adjustment at the customer level, or even device-level [10]. This kind of idea could enable the local energy system to better adjust itself when renewable energy output is volatile and small customers behave in a highly stochastic manner. Meanwhile, the robustness of energy community and distributed energy resources collective would also be improved significantly in line with the deregulated open energy market environment.

Current research on internal pricing mechanisms in the energy community includes auction-based pricing schemes and unified pricing mechanisms. Paudel et al. [11] considered demand response and proposed a novel game theory model for peer-to-peer (P2P) energy trading among community prosumers. A double auction is applied to develop a P2P energy exchange framework to enable prosumers to participate in energy sharing [12], [13]. These auction-based pricing schemes are computationally intensive and may affect transaction efficiency. The work in [14] proposes two unified internal pricing mechanisms, namely Bill Sharing and mid-market rate (MMR). Liu et al. [15] designed a pricing mechanism based on the supply and demand relationship of prosumers within the energy community. These unified pricing mechanisms are determined based on the net demand of prosumers and the external prices of grid.

Technically, for energy management of distributed resources, most of current research are inclined to adopt model-based control paradigms, including model predictive control [16], Lagrange relaxation [17], optimization algorithm [18], [19] [20], and game-theoretic solutions [21], [22]. However, it is difficult to form an accurate physical model and efficiently control the demand-side resources of the distribution network. Also, these methods will bring about high computational complexity and few intelligence features. In addition to model-based control methods, some scholars currently use cutting-edge artificial intelligence (AI) technology to solve energy management problems [23]. Reinforcement learning (RL) or deep reinforcement learning (DRL) constitutes a model-free and data-driven control framework, which gradually learns optimal control strategies through continuous interaction with the environment to accumulate experience without prior knowledge. The work in [24] established a two-layer framework for the household energy management problem and used the Markov decision process (MDP) to model the framework and applied the *rainbow* algorithm to realize economical and efficient scheduling of energy. The work in [25] designed a DRL method based on a binary deep Q-network structure for interruptible loads, and optimized the demand response management while taking into account the time-based electricity price and variable electricity consumption patterns. Wu et al. [26] proposed a hierarchical multiagent algorithm to solve the energy management problem among multiple microgrids, splitting the optimal policy learning-workload into a sequence of two subtasks. Thus, double deep Q-network

for flexible demand scheduling and deep deterministic policy gradient (DDPG) for energy trading are able to reduce the computational complexity. To effectively address the challenges of discrete-continuous mixed action spaces from home energy management, a hybrid multiagent DRL algorithm is proposed, which integrates the deep Q-learning algorithm and the DDPG algorithm [27]. In [28], a multiobjective deep reinforcement learning-home energy management system (DRL-HEMS) algorithm is proposed to optimize the operation of household devices, considering the comfort of residents as well as the load level and electricity cost of distribution transformers. However, most studies classify loads or flexible energy resources into abstract types, but not model specific load or resource types, so the working conditions of specific flexible energy resources cannot be further mapped to operation and trading state status directly [29], [30]. At the same time, most of the existing research only focuses on the possibility of theoretical research through simulation in a pure software environment, and does not explore the application of Internet-of-Things (IoT) and artificial intelligence-of-things (AIoT) [31], [32] for hardware implementation, which implies that customers do not know whether these schemes are suitable for practical implementation or not. AIoT is a relatively new technology that combines AI and IoT to enhance data management and analytics on the device at the edge [33], [34]. In the past, AI applications were mostly executed in the cloud due to the complexity of machine learning models. But for household energy management problems, control commands need to be calculated in a short period of time by running the model on user-level device based on real-time data. Therefore, the realization of AI algorithms at hardware level with limited computational resource is worth studying. This article creates a cosimulation environment using microchip-like computers and better reflects the multiagent interaction relationship of local energy management and pricing. The comprehensive simulation and experimental study could demonstrate a significant improvement in economic benefit and decision-making efficiency.

The contribution of this article is to:

1) establish various types of electrical device models, and describe their state transitions as MDP with concise representation;
2) propose an intelligent home energy management strategy for residential customers based on cutting-edge DRL algorithm with carefully designed internal pricing mechanism;
3) explore a multiagent framework for prosumers in energy community and modify the energy management models for prosumers with efficient decision-making features and local energy trading strategy;
4) conduct coordinated software and hardware cosimulation experiments for multiagent intelligent strategy based on a cost-effective IoT platform.

## II. System Framework

In this article, the overall intelligent home energy management framework contains the grid company, community energy management system (CEMS), multiple residential customers, and IoT control terminals. The system framework is shown in Fig. 1. The internal electric loads of residential customers include flexible resources with potential for regulation, such as air conditioners (AC), electric water heaters (EWH), electric vehicles (EV), and distributed energy storage. Some users also have access to distributed photovoltaic (PV). The residential customer is, therefore, considered as a prosumer with considerable power generation capability. CEMS can interact electricity power and price signals with prosumers. CEMS is also responsible for generating the local market internal electricity price according to the external electricity price assigned by the power grid company and the energy demand of prosumers. Then, the IoT control terminal transmits control signals to prosumers to regulate the corresponding electric device using predefined intelligent decision-making methods.

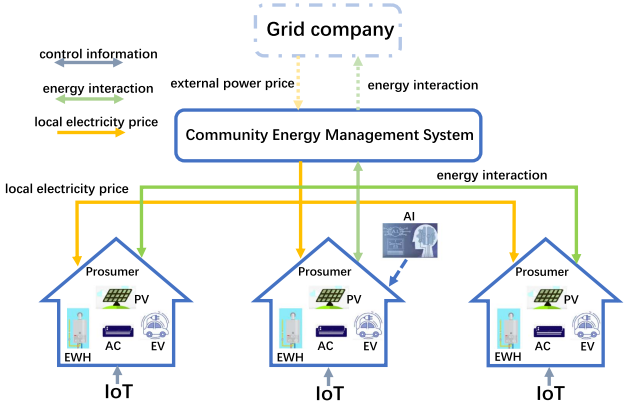


Fig. 1. Framework of intelligent home energy management based on AIoT.

The system architecture proposed in this article takes the prosumer as an agent and uses AI technology and IoT technology or combined as AIoT to manage the energy status of the prosumer. The IoT control terminal regulates the states of the electric device based on market electricity price and physical constraints of available devices. The broadcasting delay of the information interaction link among CEMS, agent and IoT terminal is limited to the millisecond level, which is ignored in this research but open to further ones. At the same time, within the user-defined time scale of 15 min, it is also assumed that the relevant states of various loads remain unchanged with constant power.

## III. Prosumer Model

Residential customers mainly include several types of electrical devices, which can be divided into adjustable loads and nonadjustable loads. The latter one mainly includes lighting, refrigerator, and other loads with low adjustment potential. For adjustable loads, three electric devices widely used in residential customers, namely AC, EWH, and EV are selected to establish the typical load consumption models for energy management.

### A. Air Conditioner

As a typical thermostatic control load, AC has great regulation potential in indoor temperature control for residential customers. The comfort of the user is ensured by controlling the indoor temperature to remain a relatively stable state.

Generally, the first-order equivalent thermal parameter model [16] is used to describe the indoor and outdoor thermal dynamic processes, and the electrical model of AC of the energy management framework in this article can be derived as

$$T^{\text{in}}_{i,t+1} = T^{\text{in}}_{i,t} e^{-\frac{\Delta t}{R_{\text{ac}} C_{\text{ac}}}} + (\eta_{\text{ac}} R_{\text{ac}} P^{\text{ac}}_{i,t} + T^{\text{out}}_{i,t})(1 - e^{-\frac{\Delta t}{R_{\text{ac}} C_{\text{ac}}}}) \tag{1}$$

$$\underline{T^{\text{in}}} \le T^{\text{in}}_{i,t} \le \overline{T^{\text{in}}} \tag{2}$$

$$P^{\text{ac}}_{i,t} = A^{\text{ac}}_{i,t} \overline{P_{i,\text{ac}}}. \tag{3}$$

In reality, the efficiency of AC is different when being operated in cooling and heating mode. It is simplified in this article, so that $\eta_{\text{ac}}$ is set as a fixed value. The range of users' indoor environment comfort is limited in (2).

### B. Electric Water Heater

EWH is selected as another typical electrical device for residential customers or prosumers. The temperature change of EWH is related to its insulation material and the use of hot water in the storage tank. This article ignores the heat loss of EWH due to the heat exchange between the insulation material and the outside, and its state transition model is as follows:

$$T^{\text{water}}_{i,t+1} = \frac{V^{\text{hot}}_{i,t} T^{\text{cold}} + (V_{\text{cistern}} - V^{\text{hot}}_{i,t}) T^{\text{water}}_{i,t}}{V_{\text{cistern}}} + \frac{C_e \eta_{\text{water}} P^{\text{water}}_{i,t}}{V_{\text{cistern}} C_{\text{water}}} \Delta t \tag{4}$$

$$\underline{T^{\text{water}}} \le T^{\text{water}}_{i,t} \le \overline{T^{\text{water}}} \tag{5}$$

$$P^{\text{water}}_{i,t} = A^{\text{water}}_{i,t} \overline{P_{i,\text{water}}}. \tag{6}$$

The thermodynamics and status transition of EWH is expressed by (4), and the user's comfortable water temperature range is limited by (6).

### C. Electric Vehicle

With its rapid development and popularization, EV has become an important electric device or flexible resource for residential customers. EV, associated with its batteries, can be considered as a typical distributed energy storage, which has characteristics similar to general energy storage. Besides, the EV model also needs to take its commuting consumption into account, by which the constraints on their charging and discharging operations are more complicated than those types of the energy storage batteries. The EV model in the CEMS for residential customers can be expressed as follows:

$$\text{SOC}_{i,t+1} = \text{SOC}_{i,t} + \frac{A^{\text{ev}}_{i,t} P^{\text{ch}}_{i,t} \eta^{\text{ch}}}{C_{\text{bat}}} \Delta t - \frac{A^{\text{ev}}_{i,t} P^{\text{dis}}_{i,t}}{\eta^{\text{dis}} C_{\text{bat}}} \Delta t - \text{SOC}^{\text{tr}}_{i,t} \tag{7}$$

$$\text{SOC}_{\min} \le \text{SOC}_{i,t} \le \text{SOC}_{\max} \tag{8}$$

$$0 \le P^{\text{ch}}_{i,t} \le \overline{P_{i,\text{ev}}}, 0 \le P^{\text{dis}}_{i,t} \le \overline{P_{i,\text{ev}}} \tag{9}$$

$$P^{\text{ch}}_{i,t} \times P^{\text{dis}}_{i,t} = 0 \tag{10}$$

$$A^{\text{ev}}_{i,t} = \begin{cases} 1 & t \notin [\alpha_{\text{ev}}, \beta_{\text{ev}}] \\ 0 & \text{else} \end{cases} \tag{11}$$

$$\text{SOC}_H \le \text{SOC}_\alpha. \tag{12}$$

The status transition of EV is expressed by (7), where $\text{SOC}^{\text{tr}}_{i,t}$ denotes the commuting energy consumption of EV at time $t$. $A^{\text{ev}}_{i,t}$ stipulates that charging and discharging cannot be carried out when EV commutes. The SOC levels at the next time bounded by the upper and lower limits and the power of charging or discharging of EV battery are, respectively, limited in (8) and (9). Constraint (10) ensures that EV cannot be charged or discharged at the same time. The limitation that EV cannot be charged or discharged during the commute period is shown in (11). The SOC of EV is required to meet basic commuting needs in (12).

### D. System Constraint of Prosumer

In addition to the aforementioned three typical loads, the prosumer resources also consist of baseline fixed loads and distributed PV. Fixed loads include electrical devices that are not easy to control, such as electric lights and refrigerators. The characteristics of fixed loads are basically constant, so the relevant data is directly given according to the actual situation without specific modeling. As a typical renewable energy resource, distributed PVs (e.g., rooftop PV panels) have developed rapidly in recent years with wide application. However, PV power generation is highly stochastic and greatly affected by solar radiation. In this article, it is assumed that the PV generation is given from day-ahead forecasting. The total system balance constraint for an individual prosumer is

$$P^{s}_{i,t} = P^{\text{fix}}_{i,t} - P^{\text{pv}}_{i,t} + P^{\text{ac}}_{i,t} + P^{\text{water}}_{i,t} + P^{\text{ch}}_{i,t} - P^{\text{dis}}_{i,t}. \tag{13}$$

### E. Local Community Internal Pricing Mechanism

In the system architecture proposed in this article, the prosumers can participate in the energy sharing community, and interact with CEMS based on real-time internal electricity price in the local market issued by CEMS. In fact, different residential customers will conduct internal energy transactions through CEMS in the community as producers or consumers in a certain period of time instead of directly trading energy with gird company. When the total net generation is less than the total net demand in the community, part of the demand is met by the local generation and the rest is supplied by grid company. The surplus energy from the community can also be sold to grid company, which plays a backup role as the last resort for energy supply. In order to make prosumers self-motivated to participate in such a local market, this article uses the MMR pricing mechanism of the P2P trading platform to give users appropriate incentives [14], [26], [35]. The local market internal pricing mechanism adopted in this article is given as follows:

$$P^{\text{producer}}_{t} = \sum_{i=1}^{N} [P^{s}_{i,t}]^{-} \tag{14}$$

$$P^{\text{consumer}}_{t} = \sum_{i=1}^{N} [P^{s}_{i,t}]^{+} \tag{15}$$

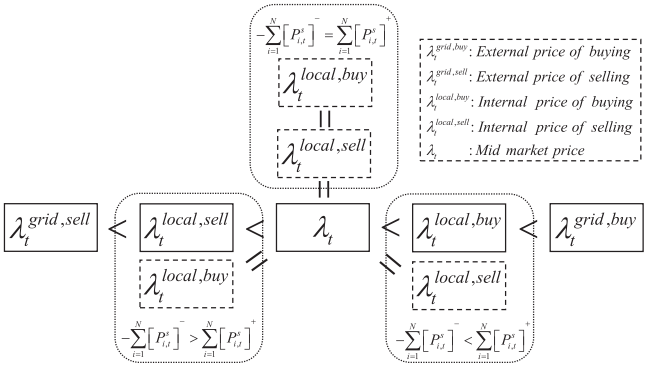


Fig. 2. MMR pricing mechanism based on the CEMS.

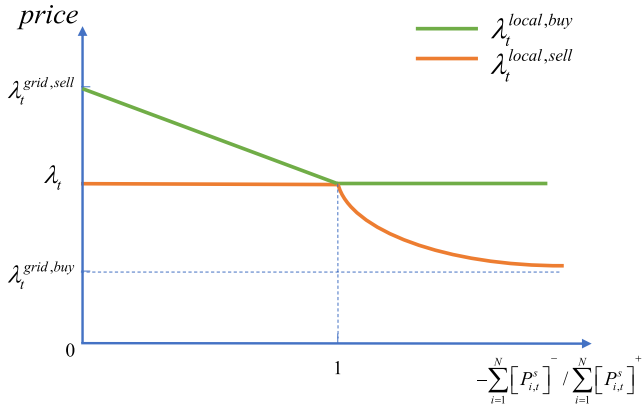


Fig. 3. Relationship between the internal prices and the supply–demand ratio.

$$P_t^{\text{total}} = P_t^{\text{producer}} + P_t^{\text{consumer}} \tag{16}$$

$$\lambda_t = (\lambda_t^{\text{grid,sell}} + \lambda_t^{\text{grid,buy}})/2 \tag{17}$$

where operators $[\cdot]^{+/-} = \max/\min\{\cdot/0\}$ indicate taking the maximum/minimum value between $\cdot$ and 0. We can divide prosumers into consumers and producers, and calculate the corresponding local market internal electricity price according to the load of the entire system in the current period.

i) When the total consumption of the community equals to the total power generation, $P_t^{\text{producer}} + P_t^{\text{consumer}} = 0$

$$\lambda_t^{\text{local,buy}} = \lambda_t^{\text{local,sell}} = \lambda_t. \tag{18}$$

ii) When the total consumption of the community exceeds the total power generation, $P_t^{\text{producer}} + P_t^{\text{consumer}} \geq 0$

$$\lambda_t^{\text{local,sell}} = \lambda_t \tag{19}$$

$$\lambda_t^{\text{local,buy}} = (\lambda_t |P_t^{\text{producer}}| + \lambda_t^{\text{grid,buy}} P_t^{\text{total}})/P_t^{\text{consumer}}. \tag{20}$$

iii) When the total consumption of the community is less than the total power generation, $P_t^{\text{producer}} + P_t^{\text{consumer}} < 0$

$$\lambda_t^{\text{local,buy}} = \lambda_t \tag{21}$$

$$\lambda_t^{\text{local,sell}} = (\lambda_t P_t^{\text{consumer}} + \lambda_t^{\text{grid,sell}} |P_t^{\text{total}}|)/|P_t^{\text{producer}}|. \tag{22}$$

As shown in Fig. 2, residential customers always sell electricity to CEMS at a price not lower than the feed-in tariff price from grid company when they are producers, and buy electricity from CEMS at a price not higher than the retail price from grid company when they are consumers. This kind of setting incentivizes residential customers to participate more in the distributed energy trading as prosumers in the local market-driven energy community.

The relationship between internal prices and the supply–demand ratio in the community is shown in Fig. 3, in which the internal prices are bounded between the feed-in tariff price and the retail price from grid company. On the other hand, this internal pricing mechanism also guarantees the economic balance in the energy sharing community.

## IV. DRL Framework

### A. Markov Decision Process

For the aforementioned three devices (AC, EWH, EV), this article assumes that the state of the next time period only depends on the current state and the control signal, and has nothing to do with the previous state, which conforms Markov property. In this article, the RL technique in AI is used to train the agent to participate in the decision-making problem of energy management. Therefore, the user's energy management problem is expressed as an MDP model, which includes the agent's state space $S$, action space $A$, reward function $R$, and state transition function $P$.

*1) State Space:* The state of the system usually changes with the agent's decision-making actions. Since the electric models keeps the Markov property, the state space of the agent can be constructed as follows:

$$S_{i,t} = [T_{i,t}^{\text{in}}, T_{i,t}^{\text{water}}, \text{SOC}_{i,t}, \lambda_t^{\text{local,sell}}, \lambda_t^{\text{local,buy}}]. \tag{23}$$

The states of prosumer $i$ at time $t$ consist of the indoor temperature, water temperature, SOC level, and the internal prices. It should be noted that in the state space, the variable of time $t$ is not included. The reason is that the influence of time on the state has been reflected in the state of the aforementioned state space.

*2) Action Space:* Actions are the results of final decisions made by the agent based on the state space, thereby controlling various types of electric devices and affecting their states. This article constructs the corresponding action space as follows:

$$a_{i,t} = [a_{i,t}^{\text{ac}}, a_{i,t}^{\text{water}}, a_{i,t}^{\text{ev}}] \tag{24}$$

$$a_{i,t}^{\text{ac}} \in [-1, 1] \tag{25}$$

$$a_{i,t}^{\text{water}} \in [0, 1] \tag{26}$$

$$a_{i,t}^{\text{ev}} \in [-1, 1] \tag{27}$$

where $a_{i,t}^{\text{ac}} \in [-1, 0)$ indicates that AC is in a cooling state; $a_{i,t}^{\text{ac}} \in (0, 1]$ indicates that AC is in a heating state. $a_{i,t}^{\text{ev}} \in [-1, 0)$ indicates that EV is in a discharging state; $a_{i,t}^{\text{ev}} \in (0, 1]$ indicates that EV is in a charging state.

*3) Reward Function:* After the agent chooses an action to execute in the action space, its state will change accordingly with a reward signal produced. As part of a feedback mechanism, the reward reflects the correctness or preference of the action chosen by the agent in the current state, thereby affecting the next decision of the agent and increasing the probability of choosing a relatively better action. The agent is trained through such a feedback loop to complete the intelligent home energy management strategy. In the intelligent decision-making model constructed in this article, the reward function is constructed as follows:

$$R_{i,t}^{\text{cost}} = -\omega_1(\lambda_t^{\text{local,sell}}[P_{i,t}^s]^- + \lambda_t^{\text{local,buy}}[P_{i,t}^s]^+) \tag{28}$$

$$R_{i,t}^{\text{ac}} = -\omega_2|T_{i,t}^{\text{in}} - T_{\text{comf}}^{\text{in}}| \tag{29}$$

$$R_{i,t}^{\text{water}} = -\omega_3|T_{i,t}^{\text{water}} - T_{\text{comf}}^{\text{water}}| \tag{30}$$

$$R_{i,t}^{\text{ev}} = \omega_4 \frac{\text{SOC}_t - \text{SOC}_{\min}}{\text{SOC}_{\max} - \text{SOC}_{\min}} \tag{31}$$

$$R_{i,t}^{s} = R_{i,t}^{\text{cost}} + R_{i,t}^{\text{ac}} + R_{i,t}^{\text{water}} + R_{i,t}^{\text{ev}} \tag{32}$$

where $\omega$ is the coefficient of each term. The reward function for buying or selling electricity is reflected in (28). Equations (29) and (30), respectively, consider the comfort of indoor temperature and water temperature. The reward function of EV takes into account the user's power anxiety in (31).

*4) State Transition:* The agent changes the state according to the actions taken. Combining the state space, action space, and the electric model constructed above, the state transition equation can be written as

$$T_{i,t+1}^{\text{in}} = T_{i,t}^{\text{in}} e^{-\frac{\Delta t}{R_{\text{ac}}C_{\text{ac}}}} + (\eta_{\text{ac}}R_{\text{ac}}a_{i,t}^{\text{ac}}\overline{P_{i,\text{ac}}} + T_{i,t}^{\text{out}})(1 - e^{-\frac{\Delta t}{R_{\text{ac}}C_{\text{ac}}}}) \tag{33}$$

$$T_{i,t+1}^{\text{water}} = \frac{V_{i,t}^{\text{hot}}T^{\text{cold}} + (V_{\text{cistern}} - V_{i,t}^{\text{hot}})T_{i,t}^{\text{water}}}{V_{\text{cistern}}} + \frac{C_e\eta_{\text{water}}a_{i,t}^{\text{water}}\overline{P_{i,\text{water}}}}{V_{\text{cistern}}C_{\text{water}}}\Delta t \tag{34}$$

$$\text{SOC}_{i,t+1} = \text{SOC}_{i,t} + \frac{a_{i,t}^{\text{ev}}\overline{P_{i,\text{ev}}}}{C_{\text{bat}}}\Delta t - \text{SOC}_{i,t}^{\text{tr}}. \tag{35}$$

The state transition equations constructed according to MDP and model of electric devices are (33)–(35). It can be seen that the state of each device of prosumer $i$ at time $t+1$ is determined by the state and action at time $t$. It should be noted that the constraints of all types of electrical device models are unchanged.

## B. DDPG Algorithm

An MDP-based model is constructed above to describe the energy management framework of residential customers. Considering the continuous characteristics of its state space and action space, it is necessary to use an RL algorithm suitable for continuous state and action spaces to train the MDP model. Deterministic policy gradient method (DPG) can solve the problem of continuous action space. DPG follows the A-C framework, and *actor* learns the policy function $\mu$ and *critic* the action value function $Q$. DPG maps the state $s$ to a deterministic action $a$ by representing the policy as a policy function $\mu_\theta(s)$. When the strategy is a deterministic one, Bellman equation is used to calculate the behavior value function $Q(s,a)$.

TABLE I
HYPERPARAMETERS OF DDPG ALGORITHM

| | | | |
|---|---|---|---|
| Step of each episode | 192 | Gamma | 0.95 |
| Learning rate of *actor* | 0.001 | BATCH SIZE | 64 |
| Learning rate of *critic* | 0.002 | MEMORY CAPACITY | 192×64 |

On the basis of DPG, the neural network is introduced to approximate the behavior value function $Q(s,a)$(*critic* network) and $\mu_\theta(s)$(*actor* network), thus forming the DDPG algorithm and realizing end-to-end learning directly from the original data. DDPG forms two independent *Target* networks for the *critic* network and the *actor* network to solve the network instability caused by frequent gradient updates.

The *actor* network chooses the action according to the current state and current policy function as

$$a_i = \mu_\theta(s_i). \tag{36}$$

Based on the Bellman equation, the *critic* network calculate the target Q-value according to current action and state from the *target* networks ($Q'$ and $\mu'$) as

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'_{i+1}|\theta^{Q'}). \tag{37}$$

The *critic* network will be updated by minimizing the loss

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2. \tag{38}$$

The gradient ascent method is used to maximize the cumulative expected reward so as to update the parameters in the *actor* network

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}. \tag{39}$$

The *target* networks will be updated through exponential moving average method as

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}. \tag{40}$$

The computation complexity of DDPG algorithm can be expressed as $O(|S| \times |A|)$, where $S$ and $A$ denote the dimension of state space and action space, respectively [36]. In this article, the DDPG algorithm is used to solve the energy management problem of residential customers, and the corresponding algorithm is particularly shown in Table I.

**Algorithm 1:** DDPG for Intelligent Home Energy Management Strategy.

**Initialize** *Critic* network $Q$ and *Actor* network $\mu$ with random weights $\theta^{Q}, \theta^{\mu}$
**Initialize** *Target* network $Q'$ and $\mu'$ with weights $\theta^{Q'}, \theta^{\mu'}$ and replay buffer $R$
**For** episode $= 1, \ldots, M$ **do**
  **Initialize** a random process $N$ for action exploration
  **Receive** initial observation state $s_1$ for prosumer
  **For** $t = 1, \ldots, T$ **do**
    Select action $a_t = [a_t^{ac}, a_t^{water}, a_t^{ev}]$ from prosumer $i$ according to the current policy by (36)
    Calculate the reward $r_t$ and the state $S_{t+1} = [T_{t+1}^{in}, T_{t+1}^{water}, SOC_{t+1}, \lambda_{t+1}^{local,sell}, \lambda_{t+1}^{local,buy}]$
    Store transition $(s_t, a_t, r_t, s_{t+1})$ in $R$
    Sample a random minibatch of $N$ transitions $(s_i, a_i, r_i, s_{i+1})$ in $R$
    Calculate $y_i$ by (37)
    Update *Critic* by minimizing the loss by (38)
    Update *Actor* network by (39)
    Update *Target* networks by (40)
  **End for**
**End for**

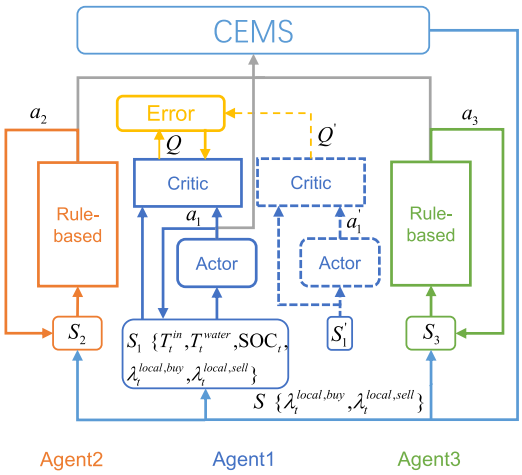


Fig. 4. Multiagent framework for simulation.

### C. Multiagent Framework

Based on the MDP model and DDPG algorithm, the intelligent home energy management decision-making problem of a single prosumer can be solved. This article studies a multiagent framework consisting of several prosumers, and the multiagent framework is shown in Fig. 4. According to the aforementioned system framework, different energy management methods are adopted for different prosumers. Agent based on DRL finds better actions by training the *actor–critic* framework, and rule-based agent selects corresponding control actions through previously established rules. A control method based on DRL is adopted for Agent 1. For the other two agents, rule-based control methods with different comfort constraints are adopted, respectively. The three agents upload the action information to CEMS and receive the real-time price issued by CEMS according to the local market internal pricing mechanism. It should be noted that the multiagent framework and energy management methods for different prosumers shown in Fig. 4 are just a schematic, and the framework can be expanded according to the complexity and actual situation.

## V. Case Study

The case study in this article takes 48 h as the horizontal timeline, assumes 15 min as the time step and, thus, trains the model across 192 time intervals each time. Some of the data used in the case study comes from realistic scenarios, including fixed load characteristic, outdoor temperature, PV output, and external electricity price [37], [38]. The rest of the data, including hourly water consumption and commuting power consumption of EV, comes from data simulations. Regarding the electricity price, this article mainly studies the energy management decision-making method of residential customers under the local market internal electricity price, which is derived from the external electricity price. Through the price mechanism setup, the peak-to-valley price difference is relatively narrowed down, and users are encouraged to conduct distributed energy transactions as prosumers.

### A. Training Process

The aforementioned MDP-based energy management model is brought into the DDPG algorithm for training. Currently, there are RL environment frameworks, such as CityLearn [39]. However, considering the independence of the model and the structure of agent, this new framework was not adopted, but we developed our own environment code based on OpenAI Gym. For the use of data, during the training process, the PV generation data and outdoor temperature data are given a 5% disturbance change for each training, thereby enhancing the reliability of the model. It should be noted that the constraints constructed in this article are not mandatory ones. In the early stage of training, it is inevitable that the state space exceeds the limit of the constraints, but the model will still continue to train till the end. These constraints constructed in this article only reflect the comfort index, and the specific constraint effect is realized by setting the reward function. The coefficient $\omega$ in the reward function is set as 1, 0.2, 0.1, and 0.5, respectively, to keep the reward of each part at the same level of magnitude. Settings of the hyperparameters of DDPG algorithm are shown in Table I.

Fig. 5 shows the changes in the reward value curves of the three agents using different regulation methods. Control methods based on different rules do not require training. After the DRL-based agent is trained over a considerable amount of episodes, the reward value curve begins to converge, which also showcases the advantages compared with the rule-based regulation methods.

For a single agent trained based on DRL (Agent 1), intelligent energy management strategy as a whole has relative advantages.

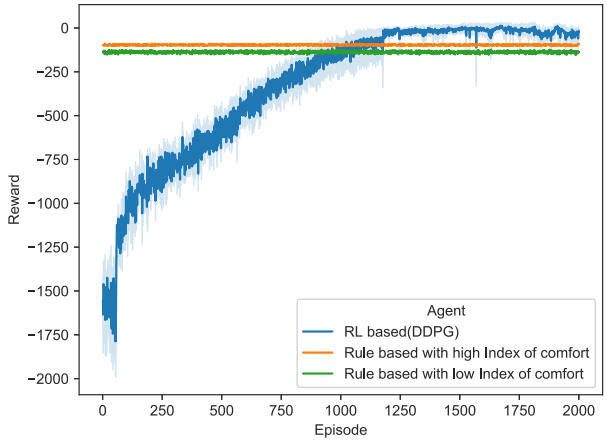


Fig. 5. Reward value curves for different agents.

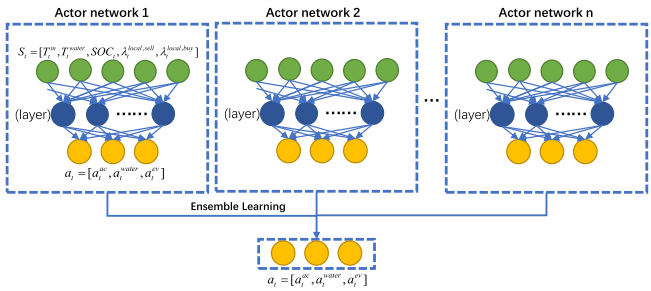


Fig. 6. Aggregate multiple actor network.

TABLE II
REWARD OF AGGREGATING MULTIPLE *ACTOR* NETWORK

| *Actor* network | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Reward | 22 | 22 | 15 | 16 |
| Reward after aggregating 2 network | 29 | | 30 | |
| Reward after aggregating 4 network | 30 | | | |

However, the action calculated by the neural network model for a single-time node is not necessarily a better choice. For Agent 1, train multiple *actor* network models and aggregate their outputs into one by using the method of ensemble learning [40]. Thus, the problem of poor regulation effect of a single-time node that may be caused by single neural network model is solved. As a whole, this kind of combination has a better effect. This logic is shown in Fig. 6, and the reward value of using the method of ensemble learning is shown in Table II.

In order to compare the effects of the proposed intelligent home energy management strategy and the rule-based control methods, the regulatory rules for the latter are defined firstly. The EV control rule is to connect EV to the charging pile at night to make up for commuting consumption, regardless of the discharge operation. When the indoor temperature exceeds the specified range or the water temperature is lower than the set value, AC and EWH will perform operations. The control rules for AC and EWH are as follows:

$$a_{i,t}^{\mathrm{ac}} = \begin{cases} 1 & T_{i,t}^{\mathrm{in}} < \underline{T^{\mathrm{in}}} \\ -1 & T_{i,t}^{\mathrm{in}} > \overline{T^{\mathrm{in}}} \\ 0 & \text{else} \end{cases} \tag{41}$$

$$a_{i,t}^{\mathrm{water}} = \begin{cases} 1 & T_{i,t}^{\mathrm{water}} < \underline{T^{\mathrm{water}}} \\ 0 & \text{else.} \end{cases} \tag{42}$$

Considering that different users have different needs for comfort, it is inappropriate to measure the effect of energy management strategies only by electricity cost. In extreme cases, if the user has no requirements for comfort, their electricity cost will be very low at this time. Hence, for the aforementioned three types of electrical devices, economic index is constructed, respectively. This economic index is not a simple electricity cost, but converts the comfort index into economic cost for consideration, specifically expressed as follows:

$$\mathrm{cost}_{i,\mathrm{ac}} = k_1 \sum_t R_{i,t}^{\mathrm{ac}} + \sum_t \lambda_t^{\mathrm{local,buy}} a_{i,t}^{\mathrm{ac}} \overline{P_{i,\mathrm{ac}}} \tag{43}$$

$$\mathrm{cost}_{i,\mathrm{water}} = k_2 \sum_t R_{i,t}^{\mathrm{water}} + \sum_t \lambda_t^{\mathrm{local,buy}} a_t^{\mathrm{water}} i, t \overline{P_{i,\mathrm{water}}} \tag{44}$$

$$\mathrm{cost}_{i,\mathrm{ev}} = k_3 \sum_t (1 - R_{i,t}^{\mathrm{ev}}) + \sum_t (\lambda_t^{\mathrm{local,buy}} a_{i,t}^{\mathrm{ev}} - \lambda_t^{\mathrm{local,sell}} a_{i,t}^{\mathrm{ev}}) \overline{P_{i,\mathrm{ev}}}. \tag{45}$$

The first part of the formula is the economic index converted from the reward function; the second part is the electricity cost of various devices. For different users, the comfort range varies. The coefficient $k$ is used to reflect the different needs of different users for comfort. In practical applications, users can set the coefficient $\omega$ in the reward function proposed before according to their own needs, so as to individualize the training process. The economic index of the DRL-based control method and the rule-based control method for various types of electrical devices are shown in Fig. 7.

### B. Implementation in IoT Platform

For the intelligent home energy management strategy for residential customers proposed in this article, the Raspberry Pi 4B is selected as the hardware device to build a software and hardware experimental platform based on IoT. As a hardware unit, Raspberry Pi 4B adopts BCM2711 processor and is equipped with quad-core ARM Cortex-A72. It runs at 1.5 GHz with 4 GB of RAM. At the software level, the 3.7.3 version of Python, the 2.2.0 version of Tensorflow and the 0.21.0 version of Gym based on the Linux system that meet the ARMv7l architecture are installed on the Raspberry Pi. As the hardware unit of the agent, the Raspberry Pi runs the DRL-based neural network model trained on the PC side and the rule-based control algorithm, respectively.

In Fig. 8, the Raspberry Pi is used as an independent IoT terminal to control the electrical devices of each residential customer. PC acts as CEMS for data communication with Raspberry Pi through WiFi. The Raspberry Pi selects the action to execute according to the corresponding state space of the residential customer and uploads its own load during this period of time to the PC through WiFi. After receiving all the information uploaded by Raspberry Pi, PC side calculates the net load within

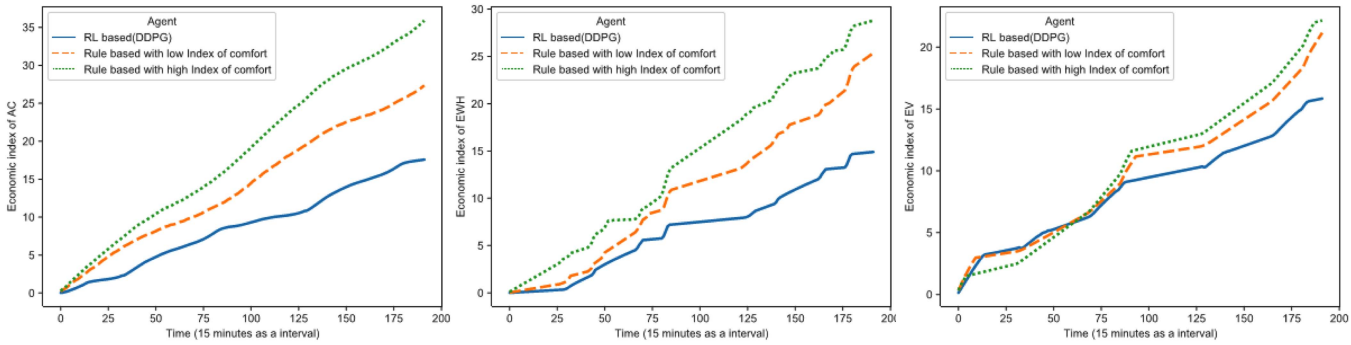


Fig. 7. Economic index of various types of electrical devices.

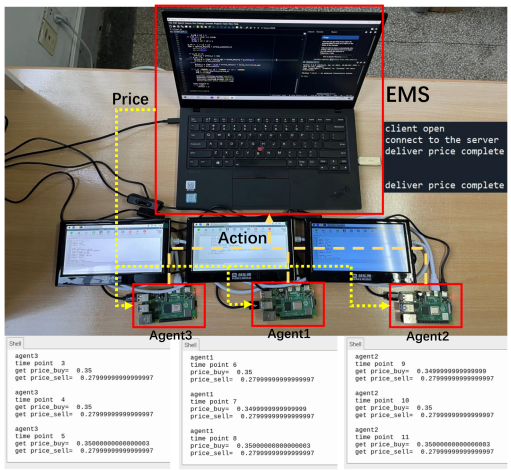


Fig. 8. Software and hardware experiment platform based on IoT.

the community and formulates the internal price within the time period according to the previously proposed internal pricing mechanism, and sends it to each Raspberry Pi. Thus, PC and Raspberry Pi together form a closed loop with the code running result shown in Fig. 8. The time scale of the energy management strategy studied in this article is 15 min. According to the actual test, the data interaction time scale between the Raspberry Pi and the PC is at the second level, so the influence of data transmission delay can be ignored.

### C. Testing of Internal Pricing Mechanism

The flexibility of various loads and PV (positive for consumption and negative for generation) is shown in Fig. 9, together with the internal price curve based on the local market internal pricing mechanism. According to the purchasing price curve of the local pricing mechanism, under the premise of following the peak and valley electricity prices, residential customers may become the role of producers during the time period of intensive PV power generation, thereby reducing the purchasing price during this time period and incentivizing residential customers to participate in distributed energy transactions. The load trend of AC is relatively flat, and the overall load shows a trend of larger loads at night and smaller loads during the day. The peak load of EWH occurs during the peak water consumption periods

TABLE III
ECONOMIC AND COMFORT INDEX OF DIFFERENT AGENT

| Agent | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Datasets | A | A | B | B | C |
| Strategy | DDPG | Rule-based | Rule-based | DDPG | DDPG |
| Electricity cost | 45.71 | 47.59 | 62.85 | 61.83 | 51.98 |
| $N_{\text{pen}}^{\text{ac}}$ | 0 | 14 | 4 | 0 | 0 |
| $N_{pen}^{ewh}$ | 0 | 9 | 23 | 2 | 0 |

of evening. Meanwhile, a higher load is also reflected in the time period of PV concentrated output and EV discharge. This phenomenon might be explained by the fact that the purchasing price of electricity has dropped at this time. EV generally charges at night and discharges during the day, which helps carrying out energy arbitrage to save costs on the premise of meeting the electricity demand for commuting.

In summary, this result verifies the effectiveness of the DRL-based multiagent energy management strategy for residential customers, and also verifies the implementation advantages of using the software and hardware cosimulation platform based on IoT technology and conceptualized AIoT framework.

### D. Extenion of Multiagent Framework

In order to increase the complexity of the model and verify the scalability of the proposed multiagent structure, two new agents are added in this case, both of which adopt the control strategy based on DDPG. In our previous work, we collected three datasets for agent training. The datasets used by each agent and the result information is shown in Table III. Similar to the training framework of three agents, the reward curve of DRL-based energy management strategy converges and shows an advantage over rule-base strategy after 2000 times of training. The rise in model complexity leads to fluctuations of the training curve. Since the agent cannot directly observe the actions and states of other agents, the participation of more agents leads to a decrease in the impact of the agent's own actions on the internal price. The information of corresponding load and internal price are shown in Fig. 10.

The change of the internal price occurs when there is surplus energy in the community, and it is bounded between the feed-in tariff price and the retail price from grid company at all times.

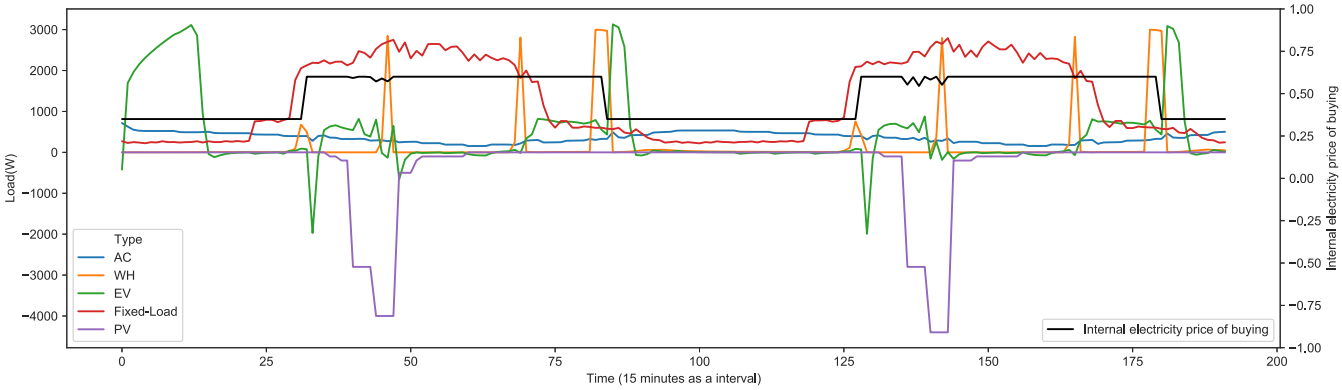


Fig. 9. Load curve of various devices and internal electricity price curve.

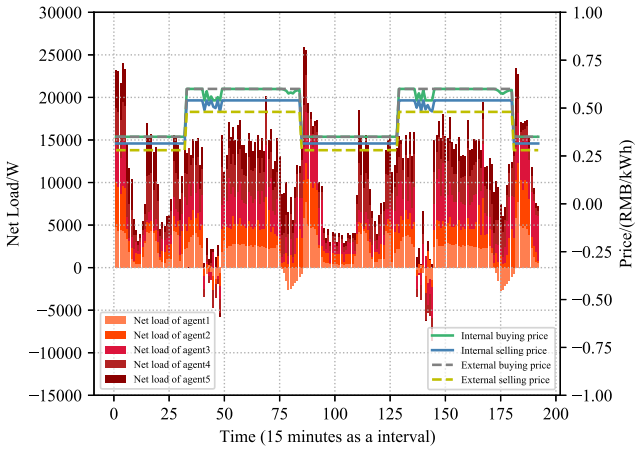


Fig. 10. Load and internal price for different agents.

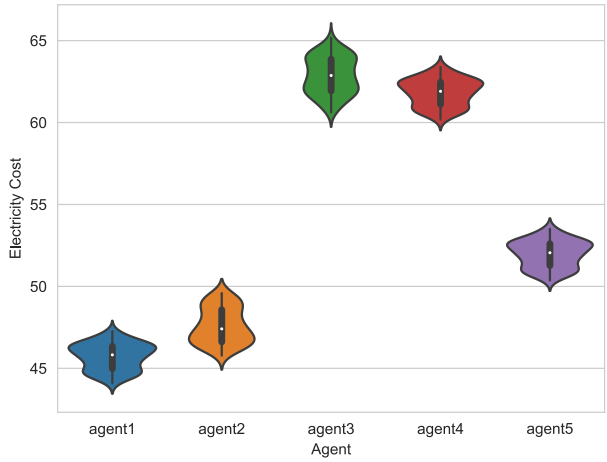


Fig. 11. Cost distribution for different agents.

Internal price signals drive load peaks in the community at night. From the perspective of economic metrics, agents 1 and 4 based on DRL reduce the cost of purchasing electricity overall according to the cost distribution of running the model for 100 times, which is shown in Fig. 11. Due to the different energy management strategies, it is obvious that rule-based regulation may save more electricity costs in terms of the insensitivity to comfort. However, if comfortable level is considered, the DRL-based energy management strategy can greatly improve user comfort while ensuring similar electricity costs.

In short, we extend the previous number of agents to five in this part of the case study. It should be noted that the case study here is only used to verify the possibility of expanding the number of agents to a few more. The generic design and expanding idea could be easily adjusted to cases that include much more agents without subtle computational consideration. It is verified that the energy management strategy based on DDPG has more advantages under the condition of incorporating the electricity cost and users' comfort at the same time.

## VI. Conclusion

In this article, a DRL-based multiagent energy management strategy is proposed for residential customers who have access to various types of distributed energy resources. In order to facilitate the local energy trading within energy community, an internal pricing mechanisms is also carefully designed. Finally, the effectiveness and advantages of the proposed methods are validated and verified on a software and hardware cosimulation platform using cost-effective IoT technology. To summarize the major findings, the conclusion can be made as follows.

1) The proposed energy management strategy can intelligently control various electric devices with self-adaptive features.
2) The automatic decision-making process can save user's electricity cost on the premise of ensuring user comfort.
3) The training effect of the DRL model can be further improved through the ensemble learning methodology.
4) The hardware and software cosimulation platform is constructed to efficiently verify the proposed intelligent energy management strategy with consideration of implementation and deployment features.

In future work, some more advanced DRL techniques and real-time data processing features can be combined with high-level energy trading strategies, which will generate another demand to upgrade the IoT hardware implementation platform as well. The distribution network featuring with more power flow consideration can be also included in the constrained MDP model with real-world deployment possibility as well.

## REFERENCES


[1] R. Lu, R. Bai, Y. Huang, Y. Li, J. Jiang, and Y. Ding, "Data-driven real-time price-based demand response for industrial facilities energy management," *Appl. Energy*, vol. 283, 2021, Art. no. 116291.

[2] T. Chen, Q. Alsafasfeh, H. Pourbabak, and W. Su, "The next-generation us retail electricity market with customers and prosumers—A bibliographical survey," *Energies*, vol. 11, no. 1, 2018, Art. no. 8.

[3] M. Song, Y. Cai, C. Gao, T. Chen, Y. Yao, and H. Ming, "Transactive energy in power distribution systems: Paving the path towards cyber-physical-social system," *Int. J. Elect. Power Energy Syst.*, vol. 142, 2022, Art. no. 108289.

[4] O. Abrishambaf, F. Lezama, P. Faria, and Z. Vale, "Towards transactive energy systems: An analysis on current trends," *Energy Strategy Rev.*, vol. 26, pp. 1–15, 2019.

[5] J. Hu, R. Harmsen, W. Crijns-Graus, E. Worrell, and V. Machteld, "Identifying barriers to large-scale integration of variable renewable electricity into the electricity market: A literature review of market design," *Renewable Sustain. Energy Rev.*, vol. 85, pp. 2181–2195, 2018.

[6] M. Song et al., "Applications of shared economy in smart grids: Shared energy storage and transactive energy," *Electricity J.*, vol. 35, no. 5, 2022, Art. no. 107128.

[7] Y. Ye, D. Qiu, X. Wu, G. Strbac, and J. Ward, "Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3068–3082, Jul. 2020.

[8] Y. Ye, D. Papadaskalopoulos, Q. Yuan, Y. Tang, and G. Strbac, "Multi-agent deep reinforcement learning for coordinated energy trading and flexibility services provision in local electricity markets," *IEEE Trans. Smart Grid*, vol. 14, no. 2, pp. 1541–1554, Mar. 2023.

[9] H. Hao, C. D. Corbin, K. Kalsi, and R. G. Pratt, "Transactive control of commercial buildings for demand response," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 774–783, Jan. 2017.

[10] N. N. Wang, "Transactive control for connected homes and neighbourhoods," *Nature Energy*, vol. 3, no. 11, pp. 907–909, 2018.

[11] A. Paudel, K. Chaudhari, C. Long, and H. B. Gooi, "Peer-to-peer energy trading in a prosumer-based community microgrid: A game-theoretic model," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6087–6097, Aug. 2019.

[12] H. Haggi and W. Sun, "Multi-round double auction-enabled peer-to-peer energy exchange in active distribution networks," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4403–4414, Sep. 2021.

[13] C. Zhang, T. Yang, and Y. Wang, "Peer-to-peer energy trading in a microgrid based on iterative double auction and blockchain," *Sustain. Energy Grids Netw.*, vol. 27, no. 6, 2021, Art. no. 100524.

[14] C. Long, J. Wu, C. Zhang, L. Thomas, M. Cheng, and N. Jenkins, "Peer-to-peer energy trading in a community microgrid," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2017, pp. 1–5.

[15] N. Liu, X. Yu, C. Wang, C. Li, L. Ma, and J. Lei, "Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3569–3583, Sep. 2017.

[16] N. Mahdavi, J. H. Braslavsky, and M. M. Seron, "Model predictive control of distributed air-conditioning loads for mitigation of solar variability," in *Proc. Australian Control Conf.*, Newcastle, NSW, Australia, 2016, pp. 162–167, doi: 10.1109/AUCC.2016.7868181.

[17] S. Bahrami, M. H. Amini, M. Shafie-Khah, and J. Catalao, "A decentralized electricity market scheme enabling demand response deployment," in *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4218–4227, Jul. 2018.

[18] Y. C. Li and S. H. Hong, "Real-time demand bidding for energy management in discrete manufacturing facilities," *IEEE Trans. Ind. Electron.*, vol. 64, no. 1, pp. 739–749, Jan. 2017.

[19] L. Yu, T. Jiang, and Y. Zou, "Online energy management for a sustainable smart home with an HVAC load and random occupancy," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1646–1659, Mar. 2019.

[20] G. Hafeez et al., "An innovative optimization strategy for efficient energy management with day-ahead demand response signal and energy consumption forecasting in smart grid using artificial neural network," *IEEE Access*, vol. 8, pp. 84415–84433, 2020.

[21] M. Yu, S. H. Hong, M. Yue, and Y. Xun, "An incentive-based demand response (DR) model considering composited DR resources," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1488–1498, Feb. 2019.

[22] B. Lokeshgupta and S. Sivasubramani, "Cooperative game theory approach for multi-objective home energy management with renewable energy integration," in *IET Smart Grid*, vol. 2, pp. 34–41, 2019.

[23] P. Yu, H. Zhang, Y. Song, H. Hui, and G. Chen, "District cooling system control for providing operating reserve based on safe deep reinforcement learning," *IEEE Trans. Power Syst.*, to be published, doi: 10.1109/TPWRS.2023.3237888.

[24] P. Du, Z. Chen, and X. Gong, "Load response potential evaluation for distribution networks: A hybrid decision-making model with intuitionistic normal cloud and unknown weight information," *Energy*, vol. 192, 2020, Art. no. 116673.

[25] B. Wang, Y. Li, W. Ming, and S. Wang, "Deep reinforcement learning method for demand response management of interruptible load," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3146–3155, Jul. 2020.

[26] Y. Wu, T. Zhao, H. Yan, M. Liu, and N. Liu, "Hierarchical hybrid multi-agent deep reinforcement learning for peer-to-peer energy trading among multiple heterogeneous microgrids," in *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2023.3250321.

[27] C. Huang, H. Zhang, L. Wang, X. Luo, and Y. Song, "Mixed deep reinforcement learning considering discrete-continuous hybrid action space for smart home energy management," *J. Modern Power Syst. Clean Energy*, vol. 10, no. 3, pp. 743–754, 2022.

[28] A. A. Amer, K. Shaban, and A. M. Massoud, "DRL-HEMS: Deep reinforcement learning agent for demand response in home energy management systems considering customers and operators perspectives," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 239–250, Jan. 2023.

[29] T. Chen and W. Su, "Indirect customer-to-customer energy trading with reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4338–4348, Jul. 2019.

[30] Q. Hu et al., "Grid-forming inverter enabled virtual power plants with inertia support capability," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 4134–4143, Sep. 2022.

[31] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet Things J.*, vol. 8, no. 10, pp. 7789–7817, May 2021.

[32] V. Puri et al., "A hybrid artificial intelligence and Internet of Things model for generation of renewable resource of energy," *IEEE Access*, vol. 7, pp. 111181–111191, 2019.

[33] I. Ghosh, "AIoT: When artificial intelligence meets the Internet of Things," 2020. Accessed: Aug. 12, 2020. [Online]. Available: https://www.visualcapitalist.com/aiot-when-ai-meets-iot-technology/

[34] Y.-J. Lin et al., "Artificial intelligence of things wearable system for cardiac disease detection," in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst.*, 2019, pp. 67–70.

[35] Y. Ye, Y. Tang, H. Wang, X.-P. Zhang, and G. Strbac, "A scalable privacy-preserving multi-agent deep reinforcement learning approach for large-scale peer-to-peer transactive energy trading," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5185–5200, Nov. 2021.

[36] G. O. Boateng, G. Sun, D. A. Mensah, D. M. Doe, R. Ou, and G. Liu, "Consortium blockchain-based spectrum trading for network slicing in 5G ran: A multi-agent deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5801–5815, Oct. 2023.

[37] "Xihe-energy," 2022. Accessed: Mar. 27, 2022. [Online]. Available: https://xihe-energy.com/#solar

[38] "China electricity price system," 2022. Accessed: Mar. 27, 2022. [Online]. Available: http://cep.cec.org.cn/api/index.html

[39] J. R. Vázquez-Canteli, J. Kämpf, G. Henze, and Z. Nagy, "CityLearn v1.0: An openai gym environment for demand response with deep reinforcement learning," in *Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Buildings Cities Transp.*, 2019, pp. 356–357.

[40] B. Xu et al., "Ensemble reinforcement learning-based supervisory control of hybrid electric vehicle for fuel economy improvement," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 2, pp. 717–727, Jun. 2020.


**Tianyun Xu** received the bachelor of engineering degree in electrical engineering, in 2022, from Chien-Shiung Wu College, Southeast University, Nanjing, China, where he is currently working toward the master of engineering graduate degree in electrical engineering with the School of Electrical Engineering, Southeast University.

His research interests include demand-side management, reinforcement learning, and zonotope technology applications in power systems.

**Tao Chen** (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Michigan-Dearborn, MI, USA, in 2018.

He is currently an Associate Professor with the School of Electrical Engineering, Southeast University, Nanjing, China. He is also affiliated with Tampere University, Tampere, Finland, working as an adjunct Researcher. Before joining Southeast University, he was a Postdoctoral Associate with Advanced Research Institute (ARI), Virginia Tech, Washington DC, USA, 2018/2019, an Intern Engineer with Global Energy Interconnection Research Institute North America (GEIRINA), Santa Clara, CA, USA, 2017/2018, and Project Researcher with the Tampere University of Technology, Tampere, Finland, 2013–2015. His research interests are about demand side management, electricity market, and machine learning applications in power systems.

**Ciwei Gao** (Senior Member, IEEE) received the M.Eng. degree in electrical engineering from Wuhan University, Wuhan, China, in 2002 and the Ph.D. degrees in electrical engineering from Shanghai Jiaotong University, China, and Polytecnico di Torino, Torino, Italy, in 2006.

He is currently a Professor with School of Electrical Engineering, Southeast University, Nanjing, China. His research interests include electricity market, demand side management, demand response, and power system planning.

**Hongxun Hui** (Member, IEEE) received the B.E. and Ph.D. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively.

From 2018 to 2019, he was a Visiting Scholar with the Advanced Research Institute, Virginia Tech, Blacksburg, VA, USA, and the CURENT Center, University of Tennessee, Knoxville, TN, USA. He is currently an Assistant Professor with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macao SAR, China. His research interests include optimization and control of power system, demand response, and Internet of Things technologies for smart energy.