IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

FDate of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Real-time Energy Management in Smart Homes through Deep Reinforcement Learning

Jamal Aldahmashi[1,2], Xiandong Ma[1]

[1]School of Engineering, Lancaster University, Lancaster LA1 4YR, UK
[2]Department of Electrical Engineering, College of Engineering, Northern Border University, Arar 73222, Saudi Arabia

Corresponding author: Xiandong Ma (xiandong.ma@lancaster.ac.uk).

The authors extend their appreciation to the Deanship of Scientific Research at the Northern Border University, Arar, KSA, for funding this research work through the project number ‘‘NBU-SAFIR-2024’’. The work was in part supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) grant EP/V040561/1.

ABSTRACT In light of the growing prevalence of distributed energy resources, energy storage systems (ESs), and electric vehicles (EVs) at the residential scale, home energy management (HEM) systems have become instrumental in amplifying economic advantages for consumers. These systems traditionally prioritize curtailing active power consumption, often at an expense of overlooking reactive power. A significant imbalance between active and reactive power can detrimentally impact the power factor in the home-to-grid interface. This research presents an innovative strategy designed to optimize the performance of HEM systems, ensuring they not only meet financial and operational goals but also enhance the power factor. The approach involves the strategic operation of flexible loads, meticulous control of thermostatic load in line with user preferences, and precise determination of active and reactive power values for both ES and EV. This optimizes cost savings and augments the power factor. Recognizing the uncertainties in user behaviors, renewable energy generations, and external temperature fluctuations, our model employs a Markov decision process for depiction. Moreover, the research advances a model-free HEM system grounded in deep reinforcement learning, thereby offering a notable proficiency in handling the multifaceted nature of smart home settings and ensuring real-time optimal load scheduling. Comprehensive assessments using real-world datasets validate our approach. Notably, the proposed methodology can elevate the power factor from 0.44 to 0.9 and achieve a significant 31.5% reduction in electricity bills, while upholding consumer satisfaction.

INDEX TERMS Power factor correction, home energy management, appliances scheduling, smart homes, reactive power compensation, deep reinforcement learning

I. INTRODUCTION

A. MOTIVATION

In today's energy research landscape, the residential sector emerges as a focal point of transformation, characterized by its accelerated melding with technological advancements, a renewed emphasis on sustainability, and an ever-growing network of interconnected systems. As detailed by seminal work in [1], three primary drivers are at the helm of this evolution: the incursion of smart control technologies, the adoption of distributed renewable resources, and the electrification of transportation via electric vehicles (EVs).

Renewable energy has always been appealing for its sustainable attributes [2]; however, it comes with a set of challenges. Due to its inherent reliance on climatic conditions, renewable energy exhibits both stochasticity and intermittency. Such unpredictable characteristics can compromise the stability of the electrical grid [3]. Concurrently, the integration of EVs introduces an additional demand, potentially establishing a new peak load. This surge might further challenge the long-term reliability and operational efficacy of the grid [4].

Energy storage (ES) solutions and flexible household appliances have garnered significant attention from both researchers and power companies, largely due to their potential in reshaping end-user consumption behaviors and overall grid dynamics based on electricity pricing [5]. With significant advancements in advanced metering infrastructure (AMI), cutting-edge sensors, and bidirectional communication channels, home energy management (HEM) systems are poised to empower consumers with proactive load monitoring and control. These systems also facilitate

real-time adjustments in consumption, on-site generation, and home ES operations contingent upon electricity prices. Such capabilities not only cater to cost efficiency by reducing electricity bills, but also fortify grid resilience, enhancing both its reliability and flexibility during critical periods [6].

Designing an effective HEM system presents a significant challenge due to the myriad uncertainties associated with end-user behaviors. These uncertainties stem not only from the inherent unpredictability of renewable energy sources but also from the dynamic nature of consumer behavior [7]. Precise forecasting of the operational duration and start time of electrical appliances for consumers remains elusive. Crafting an optimal load control strategy that satisfies both consumer preferences and network operator requirements is intricate, given the high-dimensional optimization problems being involved. Real-time response becomes even more critical when we consider that electricity, as a commodity, must be utilized immediately upon production. Hence, any household appliance scheduling that is not executed in real-time essentially renders the planning moot [8].

B. LITERATURE REVIEW

In the domain of HEM researches, a clear dichotomy emerges, differentiating approaches into two primary classifications: model-based HEM and model-free HEM. This delineation serves as a cornerstone for the formulation of subsequent energy management tactics [9].

Model-based HEM stands out as a core methodology. This can be further divided into deterministic and stochastic approaches. The deterministic model-based HEM, as referenced in sources [10-13], primarily focuses on deterministic energy cost minimization. Its main objective is the creation and subsequent resolution of optimization problems designed to determine the optimal day-ahead schedule for a variety of end-user appliances. While this method offers a strong theoretical framework, it necessitates an in-depth understanding of both operational models and detailed appliance parameters. Central to its success is the accuracy of forecasts for several external variables. Foremost among these variables are the fluctuations in utility pricing and the unpredictability of weather conditions, which significantly influence photovoltaic (PV) generation. Achieving precision in these forecasts, especially in real-time, is challenging, often leading to skepticism regarding the practical viability of this approach [14].

To address these prevailing uncertainties, the research community has increasingly turned to the use of probabilistic forecasting models as preferred predictors. These models, built upon extensive historical data and deduced parameters, aim to provide a degree of predictability within a fundamentally stochastic environment. Subsequent to the estimation process, the primary challenge emerges in the realm of control, typically addressed by employing advanced scheduling optimizers. A prominent theme in current literature is the widespread application of model predictive control (MPC) within the model-based paradigm, as highlighted in [15 - 17]. Central to the MPC algorithm is its capacity for iterative optimization, leveraging a predictive model over a dynamic time horizon.

However, the success of such model-based methodologies significantly depends on specialized knowledge. It remains incumbent upon experts to carefully design models that accurately mirror real-world dynamics, while maintaining the integrity of appliance parameters. This precision becomes even more critical when working with probabilistic predictors. Despite their potential, these predictors confront issues like determining the exact probability distribution of variable parameters, and more pressingly, computational constraints that could hinder real-time implementation [18].

A marked shift in HEM is observed towards model-free methodologies, with a particular emphasis on deep reinforcement learning (DRL) approaches [19]. The foundational strength of DRL lies in its ability to harness deep neural networks (DNN) as reliable function approximators. These DNNs, sophisticated in design and function, possess the unparalleled capability to comprehend continuous state-action transitions even when operating under ambiguous and uncertain environments [20].

When delving deeper into the intricacies of these neural architectures, it becomes evident that they are adept at accommodating and processing continuous, high-dimensional state spaces. This adaptability ensures that they are capable of discerning and extracting concealed or otherwise non-obvious attributes embedded within these state spaces. As a result, the DRL agent emerges as a robust entity, armed with the requisite capabilities to effectively address and navigate the twin challenges of environmental uncertainty and partial observability [21]. The era of the internet of things (IoT) and pervasive sensing has ushered in a data-rich landscape [22]. DRL agents, with their inherent design advantages, are ideally positioned to leverage this avalanche of data. Especially with the proliferation of intelligent sensors, these agents can consistently collect, analyze, and understand extensive datasets. This iterative and instantaneous data processing leads to the development and refinement of HEM strategies. After numerous data-driven iterations, these strategies are distinguished by their adaptability, resilience, and a remarkable capacity to succeed in fluctuating environmental conditions, especially when informed by real-time data [23]. Leveraging DRL's adaptability and learning capabilities overcomes the limitations of the traditional model-based methods, outperforming them in dynamic, uncertain environments to efficiently and effectively address HEM challenges.

The deep Q network (DQN) method stands out as a predominant approach in model-free HEM. Due to its robustness, DQN finds extensive applications across various domains including demand response management for flexible household appliances [24-25], EVs [26-27], ES systems [28-29], and heating, ventilation and air conditioning (HVAC) systems [30-31]. This is a method characterized by its proficiency in grappling with multi-dimensional continuous state spaces. However, every

technology has its limits, and DQN is no exception. It grapples with inefficiencies when deployed in continuous action domains, largely because the inherent design of the DNN is skewed towards generating discrete Q-value estimates, rendering it sub-optimal for continuous action outputs [32].

Recognizing the limitations inherent to DQN, the research community has redoubled efforts, igniting a renaissance in DRL research focused on continuous control. This resurgence has seen the advent of innovative methods, notably the deep deterministic policy gradient (DDPG), twin delayed deep deterministic policy gradient (TD3), and the advantage actor-critic (A2C). These methodologies, rigorous in design and application, have been judiciously applied across diverse platforms, from appliance scheduling and EVs to ES systems and HVAC systems [33-36]. Preliminary results, compared against traditional DQN methodologies, underscore their superior performance, offering promising avenues for future exploration in the HEM domain.

There is also a predominant focus on active power control in the HEM research, especially in the context of appliance and ES unit scheduling within residential environments. This has emerged as a dominant trend, largely due to the billing practices of most utility companies. To break it down, homeowners are predominantly billed based on their active energy consumption, with reactive power, an equally significant component of power management, often being sidelined [37]. Such a trend is not arbitrary. It is influenced by a confluence of factors, including the embryonic stage of the market for smart appliances and converters. Further, the relatively slow proliferation of HEM compounds the issue, indicating that the infrastructure and market incentives might not yet align with the growing needs of modern energy consumption [38].

The current data on energy frameworks indicates that reactive power accounts for approximately 15% to over 40% of a household's total energy consumption [39-40]. With the increasing adoption of renewable energy sources, ES systems, and EVs, this scenario presents a notable challenge. However, most of existing technologies are designed primarily to address active power demands, which directly influence monthly electricity bills, resulting in a significant decrease in active power consumption in households that utilize them. Without a corresponding reduction in reactive power consumption, a significant imbalance occurs. This imbalance can lead to an alarmingly low power factor at the interface between the home and the electrical grid, with values reaching as low as zero in some cases [41].

With the emergence of technological advancements such as smart meters, utility companies now have the capability to scrutinize both active and reactive power consumption of homeowners in greater detail. This not only provides them with enriched data but also raises the specter of financial repercussions. Specifically, homeowners demonstrating consistently low power factors might be subjected to financial penalties [42]. However, despite these progressive monitoring capabilities, a conspicuous gap persists in the scholarly examination of reactive power within residential settings.

There is a paucity of research focusing on reactive power within residential settings. While the current gap is substantial, it is anticipated to narrow progressively as HEMs become more prevalent. In studies [41], [43], a two-levels model-based optimization was implemented for the proficient management of smart converters in both ES units and EVs. The primary phase of optimization sought to minimize electricity costs, while the subsequent phase aimed to enhance the power factor. However, the practical application of this approach faces challenges. It is noteworthy that the studies incorporated a synthetic constraint by presuming the pre-availability of knowledge regarding consumer behavior and weather conditions. Furthermore, the bifurcated optimization process can potentially yield infeasible outcomes, given that solutions derived from the initial phase may not always align with the requirements of the second phase [44].

The current body of research focusing on addressing optimal HEM problems can be categorized into two distinct classifications. Table 1 presents a comprehensive overview of their respective contributions, optimization methods, reactive power control strategies, as well as their inherent limitations. Notably, an analysis of the table reveals a notable gap in the literature: to date, there has been no study that simultaneously optimizes both active and reactive power in smart homes in a real-time context.

C. CONTRIBUTION AND PAPER ORGANIZATION

In addressing the challenges encountered in conventional HEM approaches, this paper presents a novel technique grounded in DRL for a stochastic model-free HEM. This innovative methodology aims to offer an effective solution to the inefficiencies observed in existing HEM paradigms. To the best of our knowledge, this work marks the inaugural application of a real-time, model-free technique tailored for the optimal active and reactive power management in HEM contexts. The design principles guiding this approach stem from a dual objective: firstly, the mitigation of electricity expenses and dissatisfaction costs, and secondly, the augmentation of the home power factor.

In illustrating the efficacy and applicability of the proposed HEM system, comprehensive case studies employing genuine system data have been undertaken. The distinctive contributions of this work can be encapsulated into three primary areas:

1) Development of an integrated optimization framework formulated as a Markov Decision Process (MDP): The study will introduce a comprehensive optimization

TABLE 1. TABLE 1: OVERVIEW OF HEM STUDIES

Ref	Category	Contributions	Optimiz-ation Method	Reactive power control strategies	Limitations
[10]	Model-based	Developed a mixed integer linear programming (MILP) model for hybrid HEM, integrating distributed generation, ESs, and EVs with vehicle-to-home capabilities, and assessed various demand response strategies like dynamic pricing and peak power limiting for smart household energy and cost management.	MILP	X	Uncertainty of the load and PV are not considered.
[13]	Model-based	Designed an innovative HEM system featuring a smart thermostat that dynamically enhances air conditioning efficiency and comfort by adapting to variables such as electricity costs, solar exposure, and occupancy levels.	MILP	X	The approach presumes prior knowledge of user preferences, neglecting the uncertainty in load considerations.
[15]	Model-based	A HEM system, designed for optimizing residential energy usage, seamlessly incorporates PV arrays, heat pumps, and plug-in EVs. Functioning in real-time, it harmonizes electricity expenses, ES longevity, and user comfort through sophisticated predictive models, thus achieving precise energy management.	MPC	X	The system struggles with forecasting errors from mismatched predictions and actual data, along with heightened computational complexity, especially due to stochastic variables in plug-in EVs, potentially impacting optimal solution finding.
[17]	Model-based	The paper presents a new HEM system for optimal scheduling of home energy resources in high rooftop PV environments. It encompasses three stages: forecasting with an ANN for variables like solar radiation and temperature, day-ahead scheduling to minimize operation costs and manage consumption peaks, and an actual operation stage using MPC for real-time adjustments.	ANN-MPC	X	The model's complexity and computational intensity are significant, and its effectiveness hinges on the precision of the forecasting stage.
[41]	Model-based	A study introduces an optimal two-stage HEM strategy, effectively linearized into a practical format, demonstrating significant economic and technical benefits, including cost reduction and improved power factor, with innovative use of EV and ES systems for energy management and reactive power compensation.	MILP	✓	Presupposes advance knowledge of both the load uncertainty and the PV system.
[43]	Model-based	A smart HEM system is employed for integrating and coordinating various home equipment and PV system, optimally scheduling appliances, and EV, and ES systems, focusing on resident convenience, and enabling power trading. This system optimally manages both cost and power factor, while also considering load factor through the application of demand response constraints.	MILP	✓	The model does not take into account the uncertainty associated with the load and the PV generation
[23]	Model-free	The paper leverages a data-driven DQN approach to enhance energy efficiency in residential settings. The system incorporates a comprehensive reward function that simultaneously considers electricity cost, the discomfort of residents, and the life loss of transformers, aiming to balance power profiles and maximize transformer utilization without compromising user comfort. The proposed model simplifies the control mechanism by employing a single agent to manage various household appliances and load types	DQN	X	The approach to modeling air conditioning is marked by its simplicity, while the optimization method is not capable of processing continuous actions.
[25]	Model-free	A novel learning system is developed, aiming to shift loads and minimize peak aggregate load. A DQN agent is designed to simultaneously reduce consumer electricity bills and system peak load demand, with the model analyzed using loads from five residential consumers.	DQN	X	Renewable energy sources and ES are not integrated into the current model.
[26]	Model-free	An advanced HEM system is introduced, utilizing DRL to schedule home appliances and integrate customer satisfaction, employing frameworks like the Kano model for EVs and precise temperature control for air conditioners. This approach demonstrates improvements in reducing electricity costs and enhancing customer satisfaction compared to previous methods.	DQN	X	The model does not incorporate renewable energy sources and ES systems.
[27]	Model-free	Solar power is prioritized in residential EV charging, using indices to measure clean energy use and user charging preferences. The approach employs DRL and real time-of-use tariffs to optimize EV charging during high solar generation periods.	DQN	X	Load management is not addressed in this paper.
[34]	Model-free	An integrated HEM system is explored, which is engaged in a demand-side management (DSM) program and controls smart home computing tasks using a Smart Home Operation Platform. The aim is to optimize the user's total expected reward by balancing various factors such as energy costs, execution delays, and DSM compliance.	DDPG	X	Renewable energy sources are not included in the consideration.
[35]	Model-free	The demand response management problem for residential households is formulated as an MDP, considering uncertainties from sources like PV, demand, and EV. Concurrently, a model-free, data-driven DRL-based	TD3	X	The modeling approach for wet appliances is characterized by its simplicity.

		strategy is developed for managing this problem, independent of precise mathematical modeling of HEM and their uncertainties.			
[36]	Model-free	In the paper, a novel home energy recommender system was developed using DRL and MDP, with direct human feedback and resident activity data being incorporated to optimize electricity consumption and minimize resident discomfort.	A2C	X	Renewable energy source and ES are not considered

framework that considers both active and reactive power management in smart homes, uniquely employing the principles of MDP. This integrated approach is novel in the field of HEM, where most existing studies focus only on active power.

2) Real-time, model-free optimization with DRL: leveraging advanced DRL techniques, specifically the proximal policy optimization (PPO) algorithm, the research will develop a model that can dynamically and efficiently manage energy consumption in real-time. This includes adapting to changing conditions such as energy demand, supply patterns, and consumer behavior, without prior knowledge of these variables.

3) Enhancement of power factor and energy efficiency: by optimizing reactive power alongside active power, the framework aims to improve the power factor at the home-to-grid interface. This will not only enhance energy efficiency but also prevent potential financial penalties associated with low power factors.

The structure of the paper is organized as follows: section II introduces the HEM model, accompanied by a detailed problem formulation. Section III presents an innovative approach by transforming the problem into an MDP and proposing DRL-based algorithms to derive an optimal policy for the HEM. Section IV features case studies that utilize real-world residential data to validate the effectiveness of the proposed model and methodology. Section V concludes the paper and outlines potential directions for future research.

II. HEM MODEL AND PROBLEM FORMULATION

Fig. 1 depicts a smart home integrated with a HEM. This residence is connected to a low-voltage grid through a smart meter. The home is outfitted with a renewable energy source (PV), complemented by an ES unit that provides additional electricity to the household. In instances of renewable energy surplus, the HEM can choose to either store the energy in the ES system or sell it directly to the grid operator.

The residence's architecture has three layers: physical, informational, and control. The physical includes PV panels, ES units, electrical loads, and an EV. The informational layer has a smart meter for two-way communication and data exchange, like ES capacity and PV output. The control layer, led by the HEM, directs operations for the ES, EV, and electrical loads.

The electrical loads in the system are divided into three groups. The first type includes fixed loads like refrigerators and lights, which cannot be controlled by the HEM system. The second group consists of price-responsive loads such as dryers; their use can be delayed for lower energy costs but must operate within set time frames and cannot be stopped once started. The third category includes thermostatically controlled loads, which are flexible and can be adjusted by the HEM for optimal comfort. The EV is considered an interruptible load with adjustable power use, also serving as a potential energy storage asset.

The HEM operates in 15-minute intervals throughout a day, totaling 96 time steps. It collects data from various sources like sensors, PV, EV and ES, temperature, and electricity rates. The HEM then controls charging and discharging of ES and EV, manages price-responsive and thermostatically controlled loads based on tariffs. Its main goals are to reduce electricity costs, maintain optimal indoor conditions, time price-responsive load operations, and improve the residence's power factor. This strategy enhances energy efficiency, cost savings, and system performance.

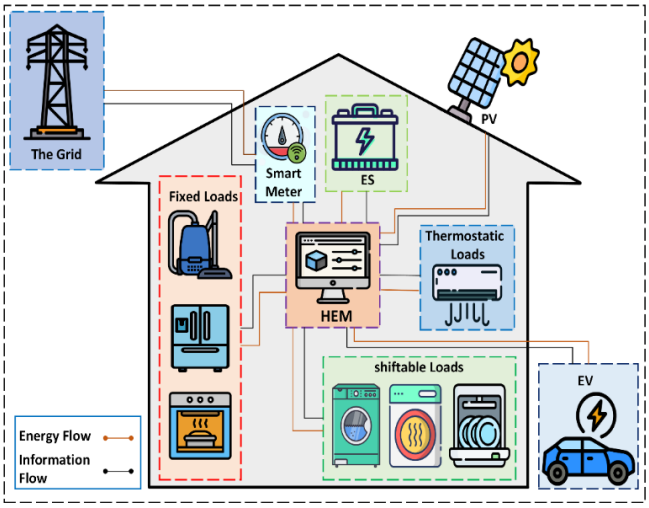


FIGURE 1. The proposed smart home architecture

A. ENERGY STORAGE UNITS

The ES unit in smart homes is key to minimizing daily electricity costs. It charges during periods of low electricity prices and discharges when prices peak. Depending on whether it's charging or discharging, the ES can act as either a stable energy source or a flexible load. Additionally, it is equipped with an advanced converter capable of producing reactive power, allowing it to operate in inductive or capacitive modes for effective power factor correction. However, the converter's limited capacity means that generating reactive power impacts the active power output of the ES unit, leading to a potential increase in electricity costs due to reduced active power production.

Given this dynamic, it becomes paramount to pinpoint the optimal balance of active and reactive power outputs from these converters. Striking this balance is pivotal for achieving dual objectives: cost efficiency on electricity bills

and proficient power factor correction. The task of managing this equilibrium rests with the HEM. Within our proposed energy management framework, the ES unit is mathematically conceptualized as follows:

$$SOE_{t+1}^{\mathrm{ES}} = SOE_t^{\mathrm{ES}} + u_t^{\mathrm{ESch}}\eta^{ESch}P_t^{\mathrm{ESch}}\Delta t - \frac{u_t^{\mathrm{ESdi}}P_t^{\mathrm{ESdi}}\Delta t}{\eta^{ESdi}}, \forall t \quad (1)$$

$$u_t^{\mathrm{ESch}} + u_t^{\mathrm{ESdi}} \le 1, \forall t \quad (2)$$

$$P_t^{\mathrm{ES}} = u_t^{\mathrm{ESch}}P_t^{\mathrm{ESch}} + u_t^{\mathrm{ESdi}}P_t^{\mathrm{ESdi}}, \forall t \quad (3)$$

$$SOE^{min,ES} \le SOE_t^{\mathrm{ES}} \le SOE^{max,ES}, \forall t \quad (4)$$

$$P^{ES,min} \le P_t^{\mathrm{ES}} \le P^{ES,max}, \forall t \quad (5)$$

$$Q^{ES,min} \le Q_t^{\mathrm{ES}} \le Q^{ES,max}, \forall t \quad (6)$$

$$S_t^{\mathrm{ES}} \le S^{ES,max}, \forall t \quad (7)$$

$$P_t^{\mathrm{ES}^2} + Q_t^{\mathrm{ES}^2} = S_t^{\mathrm{ES}^2}, \forall t \quad (8)$$

Equation (1) delineates the dynamic characteristics of the ES, wherein SOE_{t+1}^{ES} signifies the forthcoming state of energy (SOE). P_t^{ESch} and P_t^{ESdi} correspond to the power during the charging and discharging phases, respectively. The efficiency metrics for these cycles are captured by η^{ESch} and η^{ESdi}. To ensure the ES does not concurrently charge and discharge during a singular time step, binary constraints in (2), symbolized by u_t^{ESch} and u_t^{ESdi}, have been instituted, where a value of 0 indicates off and 1 denotes on. These coefficients represent the charging and discharging states, respectively. The cumulative active power involved in the charging and discharging operations of the ES is articulated in (3). Equations 4 and 5 denote the upper and lower bounds for the SOE and the active power during both charging and discharging processes. As previously highlighted, the ES converter is also equipped to engage in reactive power (Q_t^{ES}) exchanges with the smart home. Consequently, the reactive power should remain within acceptable limits as defined by (6), and the apparent power (S_t^{ES}) must not surpass the converter's rated capacity as stipulated in (7-8).

B. ELECTRIC VEHICLE

An EV serves as a versatile load, modifiable and manageable through a HEM system. The control mechanism for the charging/discharging process of the EV is a key part of this integration. The HEM system strategically manages the EV's battery charging and discharging based on real-time data, including energy demands of the household and grid conditions. This ensures optimal use of the EV's energy storage capability. The EV also offers the capability to serve as an energy source through vehicle-to-grid (V2G) or vehicle-to-home (V2H) mechanisms, thereby offering homeowners potential savings by intelligently discharging stored energy when it is most beneficial, such as during peak energy demand periods or when grid electricity prices are high. Additionally, the EV is integrated with a smart converter, adept at generating reactive power to enhance the power factor. While the EV bears resemblance to previously modelled ES systems, it exhibits additional operational constraints and functionalities, specifically in its interactive role with the HEM, where it acts not just as a load, but also as an active source participant in home energy management and grid support. For our modelling purposes, we have presupposed a singular arrival and departure time for the EV within the scheduling framework, designated as arrival time (T_{start}^{EV}) and departure time (T_{end}^{EV}). The mathematical equations representing the EV's functionality are detailed below:

$$SOE_{t+1}^{\mathrm{EV}} = SOE_t^{\mathrm{EV}} + u_t^{\mathrm{EVch}}\eta^{\mathrm{EV}ch}P_t^{\mathrm{EVch}}\Delta t - \frac{u_t^{\mathrm{EVdi}}P_t^{\mathrm{EVdi}}\Delta t}{\eta^{\mathrm{EV}di}}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (9)$$

$$u_t^{\mathrm{EVch}} + u_t^{\mathrm{EVdi}} \le 1, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (10)$$

$$P_t^{\mathrm{EV}} = u_t^{\mathrm{EVch}}P_t^{\mathrm{EVch}} + u_t^{\mathrm{EVdi}}P_t^{\mathrm{EVdi}}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (11)$$

$$SOE^{min,EV} \le SOE_t^{\mathrm{EV}} \le SOE^{max,EV}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (12)$$

$$P^{EV,min} \le P_t^{\mathrm{EV}} \le P^{EV,max}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (13)$$

$$Q^{EV,min} \le Q_t^{\mathrm{EV}} \le Q^{EV,max}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (14)$$

$$P_t^{\mathrm{EV}^2} + Q_t^{\mathrm{EV}^2} = S_t^{\mathrm{EV}^2}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (15)$$

$$S_t^{\mathrm{EV}} \le S^{EV,max}, \forall t \in \left[T_{start}^{\mathrm{EV}}, T_{end}^{\mathrm{EV}}\right] \quad (16)$$

$$SOC_t^{\mathrm{EV}} \ge SOC_{tr}^{\mathrm{EV}}, t = T_{end}^{\mathrm{EV}} \quad (17)$$

where SOE_{t+1}^{EV} denotes the SOE of the EV. P_t^{EVch} and P_t^{EVdi} symbolize the active power for charging and discharging, respectively. The efficiencies associated with these processes are represented by η^{EVch} and η^{EVdi} for charging and discharging, respectively. For effective management of the charging and discharging cycles, u_t^{EVch} and u_t^{EVdi} are introduced as binary variables indicating the respective states. P_t^{EV}, Q_t^{EV}, and S_t^{EV} collectively represent the total active, reactive, and apparent power, respectively. Equation (17) ensures that the EV maintains a sufficient charge at departure to satisfy the user's commuting needs, where SOC_{tr}^{EV} represents the percentage of energy required for the EV's travel purposes. The modeling of the EV and the ES was conducted based on the equations outlined in [35]. Additional constraints were integrated into these models to align them with the more comprehensive objectives of the paper, ensuring compatibility and relevance to the study's broader goals.

C. PRICE-RESPONSIVE LOADS

Price-responsive loads i (shiftable) operate according to user-set completion times. The HEM system schedules them

within their operating windows $[T_{start}^{\text{shift}}, T_{end}^{\text{shift}}]$, considering electricity costs and user convenience. While flexible in scheduling, these loads, once started, must run uninterrupted, hence termed "shiftable and uninterruptible loads". Examples include washing machines (WM), dryers (DM), and dishwashers (DW).

$$u_t^{shift,i} \in \{0,1\}, \forall t \in [T_{start}^{shift,i}, T_{end}^{shift,i}] \quad (18)$$

$$P_t^{shift,i} = u_t^{shift,i} U^{shift,i}, \forall t \quad (19)$$

P_t^{shift} denotes the energy consumption of flexible loads at each timestep. $u_t^{\text{shift,i}}$ represents a binary decision variable that determines the operation status of the appliance. $U^{\text{shift,i}}$ corresponds to the rated active power. Given the assumption of uninterruptibility, additional operational constraints apply to the decision variable.

$$u_t^{shift,i} = 1, \quad if \quad t \in [T_{start}^{shift,i}, T_{end}^{shift,i} - T_{require}^{shift,i}] \quad (20a)$$

$$u_t^{shift,i} = 1, if \sum_{\tau=T_{start}^{shift,i}}^{t} u_\tau^{shift,i} \leq T_{require}^{shift,i} \quad (20b)$$

$T_{\text{require}}^{\text{shift,i}}$ represents the time slots needed to fulfill the energy demand of the shiftable appliance *i*. Equation (20a) ensures the energy demand is met within the designated operating window, while (20b) guarantees uninterrupted operation of the shiftable appliance.

D. FIXED LOADS

Appliances like refrigerators and cooking devices are classified as fixed load (nonshift). Their operation is inflexible, prohibiting any scheduling adjustments. The cumulative power consumption for these appliances *j* can be represented as:

$$u_t^{nonshift,j} \in \{0,1\}, \forall t \in [T_{start}^{nonshift,j}, T_{start}^{nonshift,j} + T_{duration}^{nonshift,j}] \quad (21)$$

$$P_t^{nonshift,j} = u_t^{nonshift,j} U^{nonshift,j}, \forall t \quad (22)$$

Equation (21) delineates the binary operational status of these specific loads $u_t^{\text{nonshift,j}}$, with a value of 0 signifying an inactive state and 1 representing an active state. The operation period for these loads, as determined by the homeowner, commences at an arbitrary time point $T_{start}^{\text{nonshift,j}}$ and extends over $T_{\text{duration}}^{\text{nonshift,j}}$ time steps. Equation (22) quantifies the power of the fixed load $P_t^{\text{nonshift,j}}$ consumption of the load for each respective time interval.

E. THERMOSTATICALLY CONTROLLED LOADS

Thermostatically controlled loads like the air conditioning (AC) are classified as elastic loads due to their inherent capacity for thermal energy conservation. Their functionality is influenced by a combination of the customer's preferences, external temperatures $(Temp_t^{\text{out}})$, and current electricity prices. These loads operate to ensure that the indoor temperature $(Temp_t^{\text{in}})$ aligns with the user's desired comfort level, though they may cease operation once the desired temperature range is achieved. The calculation for the indoor temperature employs a linear equation [45], which factors in the heat exchange between the building's interior and the external environment.

$$P^{AC,min} \leq P_t^{\text{AC}} \leq P^{AC,max}, \forall t \quad (23)$$

$$Temp^{in,min} \leq Temp_t^{\text{in}} \leq Temp^{in,max}, \forall t \quad (24)$$

$$Temp_{t+1}^{\text{in}} = Temp_t^{\text{in}} - \frac{(Temp_t^{\text{in}} - Temp_t^{\text{out}} + \eta^{AC} R^{AC} P_t^{\text{AC}})\Delta t}{C^{AC} R^{AC}} \quad (25)$$

Equation (23) delineates the upper and lower bounds of energy consumption by the AC, denoted as P_t^{AC}. Equation (24) articulates the maximum and minimum thresholds for optimal internal temperature. Equation (25) defines the linear relationship governing internal temperature fluctuations, where C^{AC}, R^{AC}, and η^{AC} represent the thermal capacity, thermal resistance, and efficiency of the AC, respectively.

F. OPTIMAL DAILY ENERGY CONSUMPTION

Minimizing daily electricity expenses is a primary objective facilitated by the utilization of energy management systems.

$$l_t = P_t^{nonshift} + P_t^{shift} + P_t^{AC} + P_t^{EV} + P_t^{ES} - P_t^{PV} \quad (26)$$

$$min \sum_{t=1}^{T} \lambda_t \, l_t \, \Delta t \quad (27)$$

The l_t in (26) has two potential values: positive, representing net demand, indicating that the smart home is buying power from the grid, and negative, signifying generation, which enables the home to sell surplus power back to the grid. λ_t also encompasses two distinct values: one corresponding to the procurement cost of electricity from the grid when l_t is positive, and another denoting the sale price of electricity to the grid when l_t is negative.

III. DEEP REINFORCEMENT LEARNING-BASED SOLUTION

In this section, the optimization problem discussed in the previous section is initially converted into an MDP, and subsequently, it is addressed using a DRL algorithm.

A. MARKOV DECISION PROCESS

The decision-making framework for real-time HEM is intrinsically dependent on historical states and task assignment choices. This framework can be efficiently modeled as an MDP with an infinite temporal horizon. An MDP is defined by a three-tuple structure $(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R}(\boldsymbol{s}, \boldsymbol{a}))$, where $\boldsymbol{S}$ represents the set of all possible states, $\boldsymbol{A}$ denotes the set of actions, and $\boldsymbol{R}(\boldsymbol{s}, \boldsymbol{a})$ corresponds to the immediate rewards. The formulation details for this MDP are presented below.

1) STATE SPACE

For each time step denoted as t, the state space consolidates the data used by the agent for strategic decision-making. This state is partitioned into controllable, exogenous, and temporal segments, as described in (28) and (29). The controllable component encompasses all environmental variables directly influenced by the agent, such as the SOE of ES and SOC of EV, denoted as SOE_t^{ES} and SOC_t^{EV}, the internal temperature of the home, its power factor (PF_t), and the percentage of the energy demand B_t^{i} met by shiftable appliances. The exogenous data comprises variables that are beyond the agent's control, including external temperature, PV generation (P_t^{PV}), and electricity selling and purchasing rates (λ_t^- and λ_t^+). The time-related element represents the environment's temporal behavioral patterns, including the current time-step and the initiation of operating windows for each shiftable load and total fixed load demand P_t^{nonshift}.

$$S_t = \begin{bmatrix} t, T_{start}^{\mathrm{shift,WM}}, T_{start}^{\mathrm{shift,DW}}, T_{start}^{\mathrm{shift,DM}}, \\ Temp_t^{\mathrm{in}}, Temp_t^{\mathrm{out}}, P_t^{PV}, SOE_t^{\mathrm{ES}}, \\ SOC_t^{\mathrm{EV}}, B_t^{\mathrm{WM}}, B_t^{\mathrm{DM}}, B_t^{\mathrm{DW}}, \\ P_t^{\mathrm{nonshift}}, \lambda_t^+, \lambda_t^-, PF_t \end{bmatrix} \tag{28}$$

$$B_t^{\mathrm{i}} = \sum_{t=1}^{T} \frac{u_t^{\mathrm{shift,i}}}{T_{\mathrm{require}}^{\mathrm{shift,i}}} \tag{29}$$

2) *ACTION SPACE*

At the specified time step based on the system's state, the action of the agent is to precisely evaluate the active and reactive power flows for charging and discharging the ES and EV. Simultaneously, the agent activates the operation of the shiftable appliances, denoted as $a_t^{\mathrm{shift,i}}$ and assesses the input power of the AC, a_t^{AC}, in proportion to its rated power, $P^{AC,max}$, as described in (30).

$$a_t = \begin{bmatrix} a_t^{\mathrm{shift,WM}}, a_t^{\mathrm{shift,DW}}, a_t^{\mathrm{shift,DM}}, a_t^{\mathrm{ES,W}}, a_t^{\mathrm{ES,Var}}, \\ a_t^{\mathrm{EV,W}}, a_t^{\mathrm{EV,Var}}, a_t^{\mathrm{AC}} \end{bmatrix} \tag{30}$$

where $a_t^{\mathrm{ES,W}}$ and $a_t^{\mathrm{EV,W}}$ denote the magnitudes of charging (positive) and discharging (negative) active power for the EV and ES, respectively. Meanwhile, $a_t^{\mathrm{ES,Var}}$ and $a_t^{\mathrm{EV,Var}}$ signify the provision of reactive power (positive) and the absorption of reactive power (negative) by the converters of the EV and ES.

3) REWARD FUNCTION

The agent's primary objective is to coordinate the operation of shiftable appliances, ES, EV, and AC to chiefly reduce the daily electricity expenses for users. This involves shifting energy demand from peak pricing intervals to times with lower electricity rates. Nonetheless, such scheduling might result in user discontent, as the timing may not align with their preferences or the requisite energy demand. Moreover, this scheduling approach can adversely affect the residence's power factor, a scenario typically undesirable for network providers. Hence, from the user's viewpoint, the total reward function comprises three segments, r_t:

$$r_t = \omega_t^{\mathrm{cost}} f_t + \omega_t^{\mathrm{com}} g_t + \omega_t^{\mathrm{PF}} h_t \tag{31}$$

$$\omega_t^{\mathrm{cost}} + \omega_t^{\mathrm{com}} + \omega_t^{\mathrm{PF}} = \omega^{Total} \tag{32}$$

$$f_t = -\lambda_t \, l_t \tag{33}$$

$$g_t = g_t^{\mathrm{shiftable}} + g_t^{\mathrm{AC}} + g_t^{\mathrm{EV}} \tag{34}$$

$$g_t^{\mathrm{shiftable}} = B_t^{\mathrm{WM}} + B_t^{\mathrm{DM}} + B_t^{\mathrm{DW}} - 3, \quad if\ t > T_{end}^{\mathrm{shift,i}} \tag{35}$$

$$g_t^{\mathrm{AC}} = \begin{cases} -\left(Temp^{in,min} - Temp_t^{\mathrm{in}}\right)^2, if\ \ Temp_t^{\mathrm{in}} < Temp^{in,min} \\ -\left(Temp_t^{\mathrm{in}} - Temp^{in,max}\right)^2, if\ \ Temp_t^{\mathrm{in}} > Temp^{in,max} \\ 0\,, \qquad otherwise \end{cases} \tag{36}$$

$$g_t^{\mathrm{EV}} = \begin{cases} -(SOC_{tr}^{\mathrm{EV}} - SOC_t^{\mathrm{EV}})^3, if\ \ SOC_t^{\mathrm{EV}} < SOC_{tr}^{\mathrm{EV}} \ and\ t = T_{end}^{\mathrm{EV}} \\ 0\,, \qquad otherwise \end{cases} \tag{37}$$

$$h_t = \begin{cases} -(PF^{min} - PF_t)^2, \ \ if\ \ PF_t < PF^{min} \\ 0\,, \qquad otherwis \end{cases} \tag{38}$$

where f_t signifies the electrical expenditure associated with load energy consumption, while g_t stands for the cost related to user dissatisfaction. h_t denotes the penalty due to power factor deviations. Meanwhile, $\omega^{Total} = 1$ embodies the residential user's prioritization, capturing the intended equilibrium among electrical costs, dissatisfaction implications, and power factor penalties. A penalty is applied when the power factor at the home-to-grid interface falls below the minimum threshold, PF^{min}, set by the network operator. Fig. 2 presents the schematic representation of the MDP.

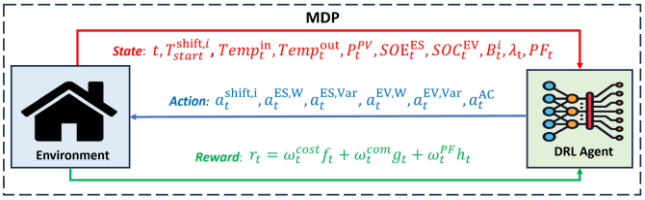


FIGURE 2. The architecture of the MDP

B. PROXIMAL POLICY OPTIMIZATION BASED SOLUTIONS

DRL represents a sophisticated approach to control where an agent operates within an MDP framework to learn and optimize policies in managing a smart home's energy system. This method enables the HEM system to effectively align with consumer goals by dynamically adjusting to changing conditions and uncertainties inherent in household operations. As shown in Fig. 3, the process involves the DRL agent constantly evaluating key home metrics, such as

electricity prices over the day and ES capacity, at each time step. Based on this evaluation, the agent makes informed decisions, receives rewards, and iteratively refines its strategies to achieve optimal energy management outcomes. Through the synergistic integration of MDP and DRL, the HEM system is thus equipped to enhance its decision-making capabilities, ensuring that household operations adapt efficiently to both the consumer's needs and the fluctuating nature of renewable energy.

The proposed energy management system is designed around PPO [51], addressing the complexities inherent in the MDP as previously detailed. Employing a model-free DRL technique, PPO functions within an actor-critic architecture, adeptly managing both continuous and discrete action spaces. Within the architecture of the PPO agent, as depicted in Fig. 3, there are two pivotal networks: the policy network (actor) and the value network (critic), each characterized by parameters, θ and α, respectively. Parameters θ and α play crucial roles in the learning process. The actor, characterized by parameters θ, is instrumental in determining the actions to be executed, mapping environmental states to respective actions. The θ parameters are integral in fine-tuning this policy to generate actions that maximize expected future rewards. Conversely, the critic, governed by parameters α, evaluates the taken actions by estimating the value of each state, providing a critique that aids in the optimization of the actor's policy. Both networks process the environmental state as input. The critic extrapolates this to yield a state value output, instrumental in fine-tuning the actor's parameters, ensuring alignment with the objective of optimizing actions. The actor, guided by this revised policy, orchestrates actions, both discrete and continuous, effectively interacting with and manipulating the environment to fulfill the desired objectives. Thus, θ and α are foundational in navigating and optimizing the decision-making process, ensuring the agent's actions are both purposeful and proficient. The actor's probability distribution is approximated by

$$\pi_\theta(a_t|s_t) = \begin{cases} \beta(p_d(s_t)) & if\ a_t \in \{a_t^{shift,WM}, a_t^{shift,DM}, a_t^{shift,DW}\} \\ N(\mu_\kappa(s_t), \sigma^2) & otherwise \end{cases} \quad (39)$$

In cases where the action is discrete, the approximation strategy follows a Bernoulli distribution $\beta(p_d(s_t))$. For continuous actions, the approximation employs a normal distribution $N(\mu_\kappa(s_t), \sigma^2)$, with $\mu_\kappa(s_t)$ and σ^2 being the mean and standard deviation of this distribution. A neural network is employed to learn these specific parameters. Based on the policy gradient technique and the gradient boosting approach, the actor's parameters, denoted as θ, are adjusted by

$$\pi_{\theta_{new}} = \pi_{\theta_{old}} + \alpha_l \nabla_{\pi_{\theta_{old}}} J(\pi_{\theta_{old}}) \quad (40)$$

where α_l represents the learning rate, $J(\pi_\theta)$ serves as the objective function for the actor, and $\nabla_{\pi_\theta} J(\pi_\theta)$ is the policy gradient. The policy gradient $J(\pi_\theta)$ is calculated using an alternative objective, referred to as $\mathbf{L^{CLIP}}(\boldsymbol{\theta})$. This surrogate objective $\mathbf{L^{CLIP}}(\boldsymbol{\theta})$ is determined by:

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[min\left(\delta t(\theta)\hat{A}_t, clip(\delta t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right)\right] \quad (41)$$

$$\delta t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \quad (42)$$

where $\hat{A}_t$ functions as the advantage function, while ε is designated as a hyperparameter and falls within the interval (0, 1). The clip function serves to truncate, setting boundaries on the variation between the prior and updated policies. Specifically, the lower and upper boundaries are $1-\varepsilon$ and $1+\varepsilon$, respectively. The advantage function can be calculated by

$$\hat{A}_t = \mathbb{E}_t[r + \gamma V(s_{t+1}) - V(s_t)] \quad (43)$$

$$V(s_t) = \mathbb{E}_\pi\left[\sum_{l=0}^{T} \gamma^l r_{t+l}\right] \quad (44)$$

where r denotes the instant reward, γ is the discounting factor, and $V(s_t)$ refers to the state value function. This function is estimated through the critic's neural network. The loss function of the state value function is defined as:

$$L^\pi(\theta) = \mathbb{E}_t[-L^{CLIP}(\theta) - c_1 H^{\pi_\theta}(s_t)] \quad (45)$$

$$H^{\pi_\theta}(s_t) = \mathbb{E}_{a_t=\pi_\theta}[\pi_\theta(a_t|s_t)\log\pi_\theta(a_t|s_t)] \quad (46)$$

where $c_1 \in [0, 1]$ is a coefficient. $H^{\pi_\theta}(s_t)$ represents the policy's entropy. Enhancing the entropy can boost the exploration capability of the PPO algorithm. For the sake of training stability, the parameters for both the actor network and the critic network are set to be shared.

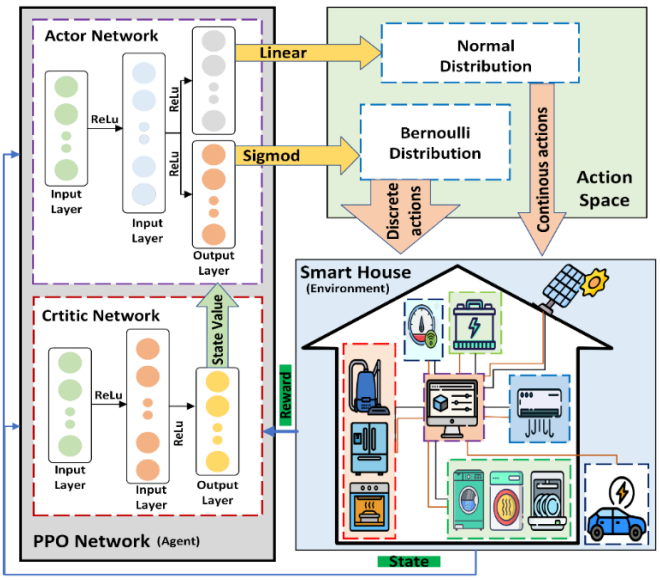


FIGURE 3. The Energy Management Algorithm based on PPO network

IV. CASE STUDY

A. EXPERIMENTAL SETUP

In this section, we assess the efficacy of the proposed HEM, which employs a model-free DRL approach, within a smart home setting. The evaluation encompasses seven fixed loads, three shiftable appliances, a thermostatic load, an ES, an EV and a PV system. The scheduling framework is based on a 24-hour period, segmented into 96 time steps, with each interval lasting 15 minutes. To introduce variability, the operating starting time for each fixed and shiftable appliance is assumed to fluctuate daily, reflecting the unpredictable nature of resident activities. Tables 2 and 3 provide an overview of the fixed and shiftable loads, respectively, and detail their starting operational hours derived from distribution sampling. Table 4 presents the technical specifications for both the ES and the EV. We posit that an EV user undertakes two journeys daily, each characterized by specific departure and arrival times. The EV's home connectivity period is anticipated to span from the conclusion of its second journey to the commencement of its first. To encapsulate the inherent uncertainties associated with consumer behavior, certain parameters, such as the EV's departure and arrival times, initial energy levels in the EV and ES ($SOE_{initial}$), and the initial indoor temperature of the residence have been modelled as random variables using a normal distribution, as elaborated in Table 5. The data presented in Tables 2, 3, and 4 have been sourced from [43].

TABLE 2. FIXED LOADS

Fixed Load	Power	PF	$T^{\text{nonshift,j}}_{\text{duration}}$	$T^{\text{nonshift,j}}_{start}$
Refrigerator	1.66 kW	0.65	96	-
Microwave	1.20 kW	0.93	1	[4-7]
			1	[46-51]
			1	[58-62]
Oven	2.4 kW	0.95	2	[2-5]
			6	[54-60]
Kettle	2.0 kW	1	1	[2-5]
			1	[46-51]
			1	[58-62]
Television	0.28 kW	0.95	20	[48-54]
Computer	0.2 kW	0.95	12	[58-62]
Lighting	0.2 kW	0.8	20	[48-54]
Vacuum	0.6 kW	0.75	2	[44-49]
Toaster	0.8 kW	0.93	1	[2-7]
Iron	2.4 kW	0.95	1	[3-6]
Security cameras	0.2 kW	0.95	96	-
Water pump	1.2 kW	0.9	3	[69-73]

TABLE 3. PRICE-RESPONSIVE LOADS

Shiftable Load	Power	PF	$T^{\text{shift,i}}_{\text{require}}$	$T^{\text{shift,i}}_{start}$	$T^{\text{shift,i}}_{end}$
Dishwasher	1.32 kW	0.7	4	[64-67]	95
Washing machine	1.4 kW	0.57	4	[44-48]	72
Clothes dryer	3.8 kW	1	4	[72-75]	95

TABLE 4. THE ES AND EV PARAMETERS

Technical parameter	EV	ES
$\text{SOE}^{\max}$ (kWh)	16 kWh	10 kWh
$\text{SOE}^{\min}$(kWh)	1.6 kWh	1 kWh
SOC^{EV}_{tr} (%)	70%	-
$\text{SOE}_{\text{initial}}$ (kWh)	[4-9] kWh	[2-6] kWh
$\text{S}^{\max}$ (kVA)	3.3 kVA	3 kVA
$P^{ch,max}_t$ (kW)	3.3 kW	3 kW
$P^{di,max}_t$ (kW)	3.3 kW	3 kW
η^{ch}, η^{di}	0.95	0.95
T^{EV}_{start}	[11-14]	-
T^{EV}_{end}	[45-50]	-

Real-world PV generation data, coupled with weather forecasting, were utilized to train and evaluate the proposed energy management system. The yearly residential PV generation data were collected from a real-world, open-source dataset gathered from households in an Australian distribution grid [46], while weather forecasting data was obtained from World Weather Online [47].

TABLE 5. THE AC PARAMETERS

Technical parameter	AC
$Temp^{in,max}$ (F)	77
$Temp^{in,min}$ (F)	68
$Temp^{\text{in,initi}}_t$(F)	[69-73]
C^{AC} (kWh/F)	0.33
R^{AC} (F/kW)	13.5
η^{AC} (kW)	2.2
$P^{AC,max}$ (kW)	1.75

Appliance simulation parameters were created using real data from a 365-day period. 300 of these days had been randomly allocated for training, with the remainder set aside for testing. It is crucial to note that the test datasets were not exposed during the training phase. Fig. 4 depicts the electricity pricing from the external grid [48]. Notably, the rate for selling electricity back to the grid was set at half the purchase price. Fig. 5 showcases a 24-hour power demand sample from the fixed loads for a residential user, beginning from 06:00 AM and concluding at the 05:45 AM slot on the subsequent day. The peak demand reached approximately 2 kW during the 21:00 time slot. From the training dataset, a random sample representing a day's PV generation for an Australian residence was selected as delineated in Fig. 6. The generation cycle commenced at 06:00 and concluded at 19:00, with the highest generation observed between 14:00 and 15:00. This PV system is characterized by a peak capacity of 1.6 kW. The PPO algorithm utilized neural networks for both its actor and critic structures. For updating the network weights, the Adam optimizer was employed, with a learning rate set at 10^{-3} for both structures. The optimization incorporates a discount factor, γ, set at 0.99. Both actor and critic networks comprise four hidden layers, each having 128 neurons. While the policy network's hidden layers used the Relu activation function, the critic value network incorporated the Tanh function. The details of the PPO algorithm's training process are presented in Algorithm 1. The DRL agent has been developed using Pytorch-2.0.1 and Python-3.10.12 on a Windows 11 system with a Core i7-12700H CPU @ 2.30 GHz × 16 and 16 GB RAM.

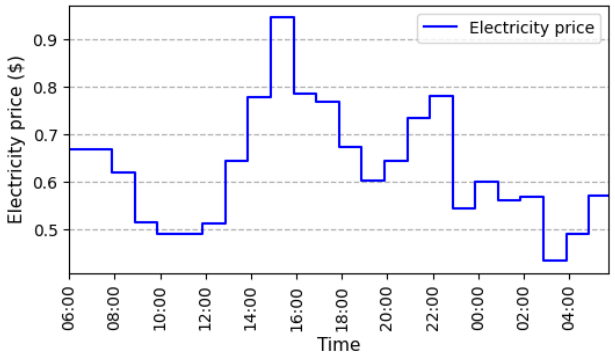


FIGURE 4. **Electricity price**

B. BENCHMARK METHODS

1) WITHOUT HEM

In this configuration, flexible appliances commence operation promptly upon task assignment. The EV undergoes charging at its peak capacity immediately upon home arrival and remains undischarged. The AC system functions at its utmost power when ($Temp_t^{\text{in}} < Temp^{in,min}$) and reverts to its minimal power setting when ($Temp_t^{\text{in}} > Temp^{in,max}$).

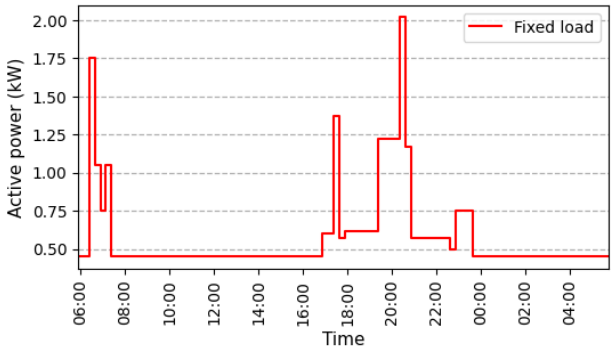


FIGURE 5. **The consumption power of fixed load**

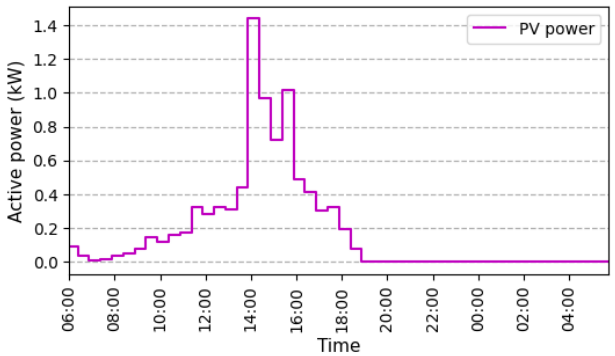


FIGURE 6. **PV output power**

2) FULL INFORMATION OBSERVABLE METHOD (FIO)

In this approach, uncertain environmental variables, including PV generation and weather data, are treated as deterministic and are presumed to be precisely known in advance. The optimization problem was systematically modeled using the equations detailed in section II, specifically by excluding the nonlinear components associated with reactive power, while focusing on active power. This approach aimed to identify the most effective appliance scheduling strategy to reduce daily electricity costs. To achieve this, the SCIP optimization toolbox [50], known for its efficiency in MILP, was employed. MILP, as a highly relevant and representative conventional method, has been extensively utilized within HEM systems. While this theoretical framework suggests an optimal boundary, it often remains unattainable in practical scenarios due to the unpredictability of environmental factors, yet it provides a critical benchmark for assessing the performance of the proposed algorithm.

3) OTHER DRL APPROACHES

The proposed PPO technique was compared with two prevalent DRL methods, specifically DQN and DDPG. To utilize DQN, which is tailored for discrete action spaces, a Q-network comprising 6 hidden layers with neuron configurations of 128,128,128, 64, 64, and 64 using ReLU activation functions is employed to estimate the Q-function. Consequently, the action space is divided into 576 distinct possibilities. Conversely, for the application of DDPG, designed exclusively for continuous action spaces, an actor network structured with 4 hidden layers consisting of 128,128,128, and 64 ReLU neurons is used to pinpoint the best action. Given that the control variables associated with shiftable loads are binary in nature, the output from the actor network must be converted into binary form. If this output falls below 0, a binary action of 0 is chosen; otherwise, it defaults to 1. A critic network, with an architecture mirroring that of the actor, is employed to estimate the optimal value function. The implementation of these two algorithms was carried out in accordance with the guidelines outlined in [52]. Adjustments were made to the number of states, actions, and DNN layers to tailor the algorithms specifically for the smart home environment proposed in this study.

Algorithm 1: Process for training the PPO agent.

Input: the state of smart home environment
Output: actor network $\pi_\theta$is employed for real-time energy management within smart
Initialize parameters θ and α randomly
Initialize old actor parameters: $\pi_{\theta_{\text{old}}} \leftarrow \pi_\theta$
For $episode = 1, 2, \ldots, E$ **do**:
  Reset the initial state of the environment randomly
  For $t = 1, 2, \ldots, T$ **do**:
    Observe the state s_t according to (28)
    Sample action a_t based on $\pi_{\theta_{\text{old}}}$
    Calculate reward r_t and obtain new state s_{t+1} according to (31)
    Store $(s_t, a_t, s_{t+1}, \log \pi_\theta(a_t|s_t))$ in memory buffer
  end for
  for $n = 1, 2, \ldots, N$ do
    Calculate $\hat{A}_t$ based on (42)
    Calculate $L^\pi(\theta)$ based on (44)
    Optimize the loss function with respect to θ
  end for
end for

C. EXPERIMENTAL RESULTS

1) CONVERGENCE PERFORMANCE AND COST REDUCTION

a: TRAINING PERFORMANCE

The PPO, DQN, and DDPG methods are trained over 5000 episodes to learn the optimal strategy for residential energy management. Fig. 7 illustrates the progression of cumulative rewards throughout the training process. Within the figure, solid curves depict the mean cumulative rewards, aggregated across five seeds, and shaded areas represent the corresponding standard deviation values. As illustrated in the figure, in the early stages of learning, average cumulative rewards are low since the agents are mainly exploring various actions without much direction. However, as training advances and the agents gain more experience, the rewards increase, peaking for all three methods. Notably, right from the onset, PPO's rewards rise more rapidly than those of DQN and DDPG. The comparative performance of various DRL algorithms, including the proposed one, is delineated in Table 6, highlighting their stabilization and convergence to near-optimal solutions. Remarkably, the proposed algorithm not only requires fewer episodes to converge compared to other algorithms but also demonstrates a shorter training duration.

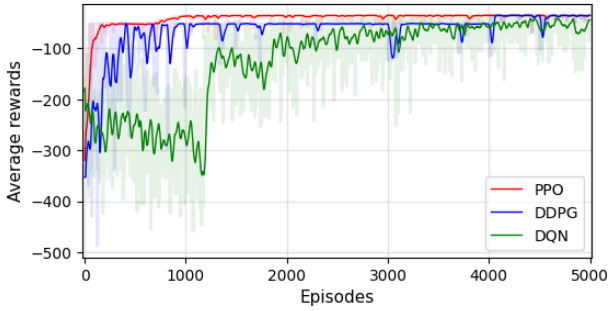


FIGURE 7. Average Reward for the PPO, DQN and DDPG methods

TABLE 6. TRAINING PERFORMANCE COMPARISON OF DRL ALGORITHMS

DRL Algorithm	Episodes to Convergence	Average Reward (r_t)	Training Time (min)
DQN	3562	-29.82	258
DDPG	4031	-28.29	294
PPO	**994**	**-28.25**	**237**

b: TESTING PERFORMANCE IN ELECTRICITY COST REDUCTION

To evaluate the effectiveness of the PPO algorithm specifically in reducing electricity expenses, the PPO underwent testing with new datasets that were not used during the initial training phase of the agent. For a comprehensive quality assessment, the proposed solution was benchmarked against four reference strategies, i.e., the without HEM policy, the FIO policy, the DQN algorithm, and the DDPG algorithm. It is imperative to mention that, while the FIO policy represents an optimal performance benchmark, its realization in practical scenarios is limited due to inherent randomness, such as unpredictable consumer behavior and volatile weather conditions.

Fig. 8 illustrates a comprehensive analysis of the total costs associated with various strategies over a span of 15 testing days. The data presented in this figure reveals that our proposed method markedly outperforms the Without HEM baseline, registering a 31.5% reduction in electricity expenses. In a comparative analysis with other DRL algorithms such as DQN and DDPG, our proposed method also exhibits superior performance. DQN and DDPG achieved a reduction in the electricity bill by 18.64% and 24.77% respectively, falling short of the efficacy demonstrated by our approach. The FIO algorithm, notable for its proficiency in a non-random environment, attains an ideal cost reduction of 39.62%. This is attributed to its tailored design that is optimally responsive to predictable, structured settings. In contrast, our proposed algorithm is engineered to adeptly navigate through random environments. Despite the inherent challenges of unpredictability and variability, it approximates the performance of the FIO algorithm closely. Table 7 presents a comprehensive evaluation of the average electricity costs incurred utilizing various optimization strategies throughout the testing days. Notably, the PPO method demonstrated the lowest cost, markedly outperforming methods tailored for stochastic environments. Moreover, the results underscore the efficiency of the PPO method in attaining a near-optimal solution within a markedly brief duration of 0.35 seconds. This performance is contrasted with the FIO method, which required a substantially longer time of 27 seconds. These findings emphasize the PPO method's capability in enhancing the management of household loads, promoting optimal energy use, and effectively minimizing costs within a real-time operational framework.

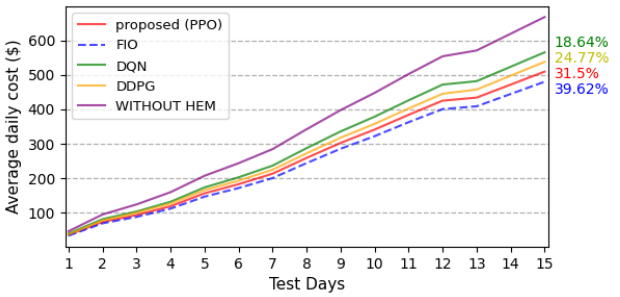


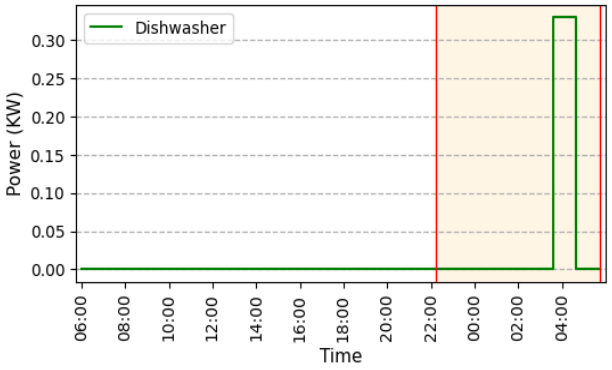
FIGURE 8. Average daily cost over 15 days

TABLE 7. COMPARATIVE ANALYSIS OF ELECTRICITY COST REDUCTION STRATEGIES

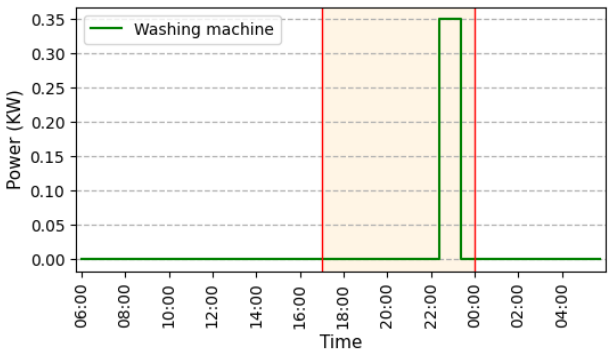
STRATEGY	AVERAGE ELECTRICITY COST ($)	RESPONSE TIME (S)
WITHOUT HEM	43.87	-
FIO	31.40	27.12
DQN	37.01	0.45
DDPG	34.67	0.41
PPO (Proposed)	**33.34**	**0.35**

2) APPLIANCE SCHEDULING EFFECTIVENESS

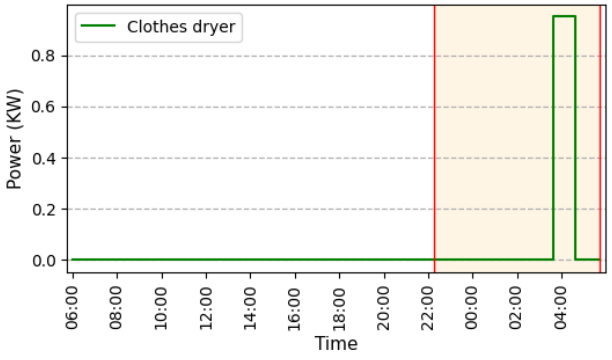
The efficacy of the proposed algorithm in load scheduling was assessed through a meticulous evaluation process. A specific day was randomly selected from the testing phase to facilitate a comprehensive analysis, the outcomes of which are illustrated in Fig. 9. The graphical representations in Fig. 9(a), (b) and (c) conspicuously demonstrate the algorithm's adeptness at scheduling flexible loads within the consumer's preferred operational timeframes. A pivotal attribute of the algorithm is its proficiency in capitalizing on periods when electricity prices are minimal, engendering both economic and energy efficiencies. Each subfigure denotes the scheduling window of the corresponding appliance, as illustrated by the highlighted orange region. A critical examination of thermostatic loads, particularly given their significant impact on consumer comfort, was also conducted.



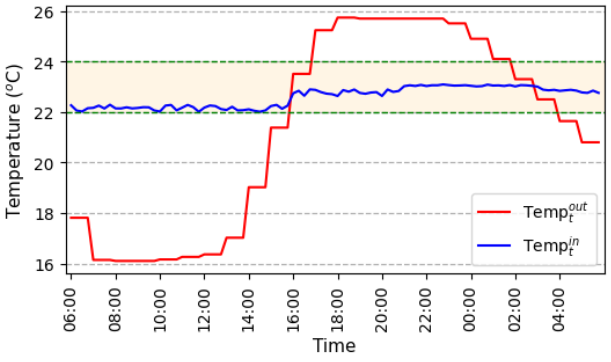
(a)



(b)



(c)



(d)

FIGURE 9. The performance of the proposed method showing the appliance scheduling effectiveness, (a) dishwasher, (b) washing machine, (c) clothes dryer, and (d) air conditioning unit

The algorithm exhibited a remarkable capability in the efficient management of AC systems. Fig. 9(d) underscores the algorithm's effectiveness in maintaining indoor temperature fluctuations within the thresholds of comfort, as shown in the orange region, ensuring an optimal ambient environment. This comprehensive assessment demonstrates the robustness and applicability of the proposed algorithm in real-world scenarios, showcasing its potential to enhance both energy efficiency and consumer comfort significantly.

3) SCHEDULE OF ES AND EV

ES and EV have the potential to optimize energy consumption through the HEM, specifically by modulating their charge and discharge cycles in response to fluctuating energy prices. However, identifying the best strategy to meet consumer objectives poses significant challenges, given the unpredictability in EV usage patterns, external temperature variations, and the inconsistent output of PV systems. Nevertheless, our proposed algorithm effectively navigates these uncertainties, primarily focusing on two outcomes: minimizing electricity costs and enhancing user experience. As illustrated in Fig. 10, the charge and discharge cycles of the active power of the ES are influenced by both electricity pricing and the productivity of PV systems. The initial state of energy, $SOE_{initial}$, for the ES was randomly set at 4 kWh. Due to the relatively high electricity prices occurring between 6 and 8 a.m., coupled with the minimal PV power production during this period, the HEM opted to discharge the ES. This approach was preferred to minimize reliance on energy from the public network. In contrast, between 8 and 10 a.m., electricity prices are lower relative to other times. Thus, the HEM favored charging the ES during this period, with a plan to discharge it subsequently. It is noted that the production of PV peaks at 2 p.m. Leveraging this, the HEM strategically decided to store this energy, intending to discharge it over the remaining part of the day as a measure to optimize energy costs effectively.

Fig. 11 illustrates the varying patterns of charging and discharging between an EV battery and an ES. This disparity is primarily attributed to the specific constraints placed on

the EV battery. A crucial limitation influencing user convenience is the necessity for the EV battery to maintain ample energy reserves, ensuring the uninterrupted operation of the EV. During the early hours, specifically from 6:00 to 7:30 a.m., when electricity rates are comparatively high, the HEM strategically opts for minimal battery charging. This approach guarantees that the EV battery receives a sufficient charge before departure, aligning with user convenience and operational readiness. As electricity costs diminish at 8 a.m., becoming more economical than earlier hours, the HEM decides to amplify the charging intensity, maximizing the battery's energy uptake. Within Fig. 11, the orange segments denote periods when the EV is stationed at home, indicating availability for charging. It is essential to note that the EV's battery should maintain a minimum of 70% charge (SOC_{tr}^{EV}) before departure from a smart home environment, as shown in Fig. 12. The HEM system treated the EV as a load before departure time, resulting in a distinct operational behavior for the EV compared to the ES system during this period. A dashed line, in Fig. 12, represents periods where the EV's battery status remains undetermined by the HEM. However, on the EV's return home at 17:30, it was discerned by the HEM that the battery retained a 31% charge. The HEM regards the EV upon its arrival as an energy source and notes that the EV's charging and discharging actions are similar to the behavior of the ES throughout this period. Consequently, leveraging periods of elevated electricity prices, specifically between 19:00 and 23:00, the energy was strategically discharged, optimizing cost-efficiency and energy utilization.

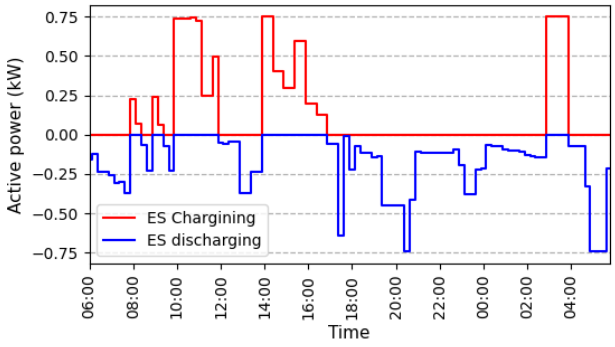


FIGURE 10. Charge and discharge cycles of ES

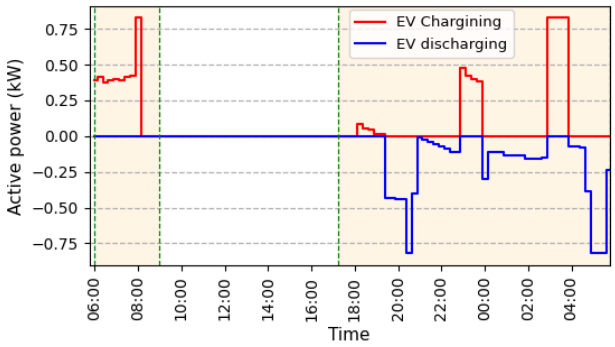


FIGURE 11. Charge and discharge cycles of EV

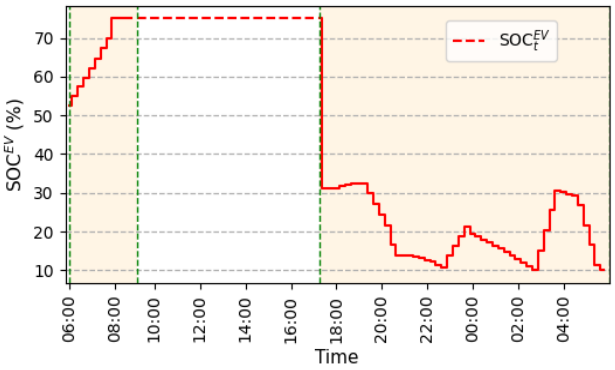


FIGURE 12. State of charge electric vehicle

4) REDUCTION OF REACTIVE POWER AND POWER FACTOR CORRECTION

HEM primarily aims to curtail electricity costs by minimizing active power consumption from the grid, particularly during peak pricing periods. A common oversight in many HEMs is the management of reactive power. This is largely because reactive power does not have a direct impact on cost reduction and involves more intricate optimization strategies compared to managing active power compensators. Fig. 13 illustrates the extent of reactive power extracted from the grid due to the utilization of a HEM based on the FIO method that omits reactive power considerations. In previous approaches, the FIO method was employed to coordinate ES and EV converter schedules for active power reductions. However, these converters were not tasked with compensating for the reactive power essential for both flexible and fixed loads. Leveraging the advanced capabilities of the proposed method, which adeptly addresses high-dimensional optimization problems, the proposed DRL agent was trained. Its objective was to minimize the reactive power import from the grid by effectively managing ES and EV converters. As depicted in Fig. 13, the proposed algorithm adeptly controls the converters, ensuring they compensate for the reactive power at their connection point to the grid. Table 8 illustrates that the application of the PPO algorithm results in an average reduction of reactive power drawn from the network and hence the increase of power factor. This reduction in reactive power is threefold when compared to the FIO algorithm. This significant distinction highlights the vital importance of reactive power control in smart homes to ensure the stability of the electrical network.

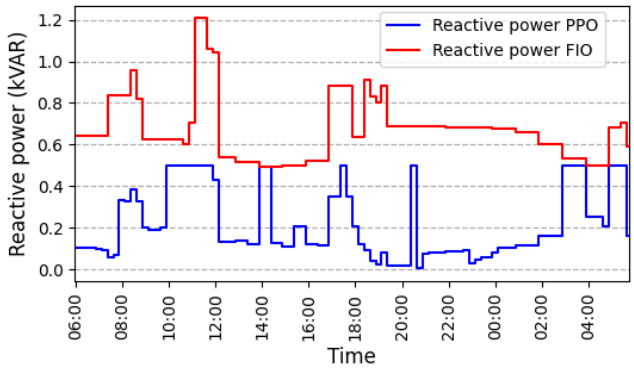


FIGURE 13. Reactive power import from the gird

TABLE 8. REACTIVE POWER REDUCTION AND POWER FACTOR CORRECTION IN SMART HOMES

STRATEGY	REACTIVE POWER (kVAR)	AVERAGE POWER FACTOR
FIO	16.01	0.44
PPO	**5.06**	**0.901**

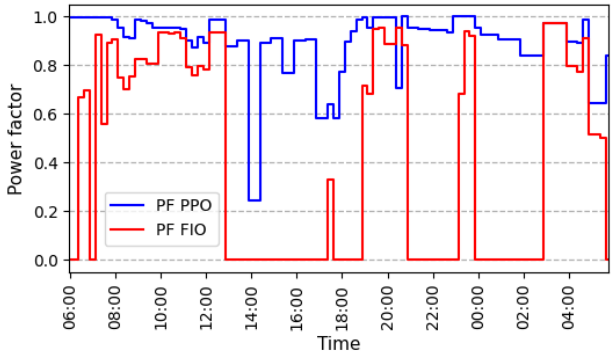


FIGURE 14. Power factor record at the home-to-grid integration point

Fig. 14 contrasts the variation in the power factor when employing a FIO against one harnessing PPO method. A discernible decline in the power factor is observed when employing the PPO method between the time intervals of 13:45 to 17:45. This decline can be attributed to the unavailability of the EV, which typically provides the requisite reactive energy to support the household loads. The absence of this energy source results in a diminished power factor during these intervals. In certain service provisions, it is mandated that the power factor for residential properties should surpass 0.9 [49]. Non-compliance, evidenced by a power factor below this threshold, leads to penalties for the property owners. Through the implementation of the proposed PPO technique, we achieved an average power factor of 0.901 for the household. In contrast, utilizing the FIO method resulted in an average power factor of 0.44, a value that is not viable for upcoming electrical grid systems.

V. CONCLUSION

In the present study, we presented a real-time model-free HEM approach grounded on the PPO algorithm. This approach is tailored to optimize the scheduling and management of a diverse range of loads, including EV and ES, within smart homes. To ascertain its efficacy in uncertain conditions, we modelled fixed, flexible, and thermostatic load, as well as EV and PV generations in a stochastic manner. Upon comparison with benchmark methods such as DQN, DDPG and FIO, our proposed approach demonstrated superior efficiency and effectiveness. When evaluated using real-world datasets, it consistently exhibited exceptional performance, particularly in the scheduling of flexible and thermal loads in alignment with the preferences of the end-user. A noteworthy outcome was the algorithm's ability to proficiently manage the active power of ES and EV. This capability significantly contributed to a substantial reduction in electricity bills by 31.5%. Moreover, the algorithm is able to adeptly control the bidirectional converters linked to both the ES and EV, optimizing the injection of necessary reactive power. This optimization enhanced the overall power factor of the smart home, elevating it to 0.901 from 0.44 resulting from the benchmark algorithm. In summary, the adoption of our proposed method promises substantial economic and operational dividends, benefiting both consumers and utilities. Future work will be centered around optimizing the power factor within smart homes during the periods when an EV is unavailable.

ACKNOWLEDGMENT

Jamal Aldahmashi extends his sincere appreciation to Northern Border University for fully funding his Ph.D. research at Lancaster University.

NOMENCLATURE

SOE^{ES}	State of energy of the ES (kWh)
u^{ESch}, u^{ESdi}	Binary variables indicating if the ES is charging / discharging process
η^{ESch}, η^{ESdi}	Efficiency of the ES system during charging and discharging process
P_t^{ESch}, P_t^{ESdi}	Power related to the ES system charging / discharging process (kW)
$SOE^{min,ES}, SOE^{max,ES}$	Minimum and maximum allowable state of energy for the ES (kWh)
$P^{ES,min}, P^{ES,max}$	Minimum and maximum power limits for the ES (kW)
$Q^{ES,min}, Q^{ES,max}$	Minimum and maximum reactive power limits for the ES (kVAR)
$S^{ES,max}$	Maximum apparent power limit for the ES (kVA)
t	Time step
Δt	Time interval
SOE^{EV}	State of energy of the EV (kWh)
u^{EVch}, u^{EVdi}	Binary variables indicating if the EV is charging / discharging process
η^{EVch}, η^{EVdi}	Efficiency of the EV system during charging and discharging process
P_t^{EVch}, P_t^{EVdi}	Power related to the EV system charging / discharging process (kW)
P_t^{PV}	Power related to the PV system at t (kW)
$SOE^{min,EV}, SOE^{max,EV}$	Minimum and maximum allowable state of energy for the EV (kWh)
$P^{EV,min}, P^{EV,max}$	Minimum and maximum power limits for the EV (kW)
$Q^{EV,min}, Q^{EV,max}$	Minimum and maximum reactive power limits for the EV (kVAR)
$T_{start}^{EV}, T_{end}^{EV}$	Arrival and departure time
SOC_t^{EV}	State of charge of an EV at t (%)
SOC_{tr}^{EV}	Minimum state of charge necessary for completing a specific trip (%)
$u_t^{shift,i}$	Binary variable indicating whether a shiftable load i is operating at t
$P_t^{shift,i}$	Power related to the shiftable load i (kW)
$U^{shift,i}$	Rated power of shiftable load i (kW)
P_t^{shift}	Total shiftable load at t (kW)
$T_{require}^{shift,i}$, $T_{start}^{shift,i}, T_{end}^{shift,i}$	Required operation, start and end, time for shiftable load i
$P_t^{nonshift}$	Total fixed load at t (kW)
$u_t^{nonshift,j}$	Binary variable indicating whether a fixed load j is operating at t
$U^{nonshift,j}$	Rated power of fixed load j (kW)

$T_{duration}^{nonshift,j}$	Operational time period of fixed load j
P_t^{AC}	Power related to the AC system (kW)
$P^{AC,min}, P^{AC,max}$	Minimum and maximum power limits for the AC system
$Temp_t^{in}$	Indoor temperature (°F)
$Temp^{in,min}, Temp^{in,max}$	Minimum and maximum indoor temperature limits (°F)
$Temp_t^{out}$	Outdoor temperature (°F)
η^{AC}	Efficiency of the AC
R^{AC}	Thermal resistance (F/kW)
C^{AC}	Thermal capacity of AC (kWh/F)
l_t	Total load at t (kW)
λ_t	Electricity price ($)
S_t	State vector
PF_t	Power factor at t
PF^{min}	Minimum Power factor
$B_t^{WM}, B_t^{DM}, B_t^{DW}$	The percentage of the energy demand met by shiftable appliances
a_t	Action vector at time t
$a_t^{shift,WM}, a_t^{shift,DW}, a_t^{shi}$	Actions related to shiftable loads
$a_t^{ES,W}, a_t^{ES,Var}$	Actions related to ES active and reactive power
$a_t^{EV,W}, a_t^{EV,Var}$	Actions related to EV active and reactive power
a_t^{AC}	Action related to AC
r_t	Reward at t
$\omega_t^{cost}, \omega_t^{com}, \omega_t^{PF}$	Weights related to electricity cost, user dissatisfaction, and PF deviation
g_t	Cost related to user dissatisfaction
h_t	Cost related to PF deviation
$\pi_\theta(a_t \mid s_t)$	Policy function parameterized by θ
$L^{CLIP}(\theta)$	The clipped surrogate objective function
$J(\pi_\theta)$	The objective function of PPO
$\hat{A}_t$	Advantage at episode t.
$V(s_t)$	Value function
$L^{\pi}(\theta)$	Policy loss function
$H^{\pi_\theta}(s_t)$	Entropy of the policy π_θ at state s_t
E	Total number of episodes
N	Number of iterations for updating the policy

REFERENCES

[1] H. Shareef, M. S. Ahmed, A. Mohamed, and E. Al Hassan, "Review on home energy management system considering demand responses, smart technologies, and intelligent controllers," IEEE Access, vol. 6, pp. 24498-24509, 2018.

[2] S. T. Meraj, N. Z. Yahaya, K. Hasan, M. H. Lipu, R. M. Elavarasan, A. Hussain, M. A. Hannan, and K. M. Muttaqi, "A filter less improved control scheme for active/reactive energy management in fuel cell integrated grid system with harmonic reduction ability," Applied Energy, vol. 312, p. 118784, 2022.

[3] S. Mokeke and L. Z. Thamae, "The impact of intermittent renewable energy generators on Lesotho national electricity grid," Electric Power Systems Research, vol. 196, p. 107196, 2021.

[4] K. Valogianni, W. Ketter, J. Collins, and D. Zhdanov, "Sustainable electric vehicle charging using adaptive pricing," Production and Operations Management, vol. 29, no. 6, pp. 1550-1572, 2020.

[5] P. Munankarmi, J. Maguire, S. P. Balamurugan, M. Blonsky, D. Roberts, and X. Jin, "Community-scale interaction of energy efficiency and demand flexibility in residential buildings," Applied Energy, vol. 298, p. 117149, 2021.

[6] H. T. Nguyen, U. Safder, J. Loy-Benitez, and C. Yoo, "Optimal demand side management scheduling-based bidirectional regulation of energy distribution network for multi-residential demand response with self-produced renewable energy," Applied Energy, vol. 322, p. 119425, 2022.

[7] C. Ziras, C. Heinrich, M. Pertl, and H. W. Bindner, "Experimental flexibility identification of aggregated residential thermal loads using behind-the-meter data," Applied Energy, vol. 242, pp. 1407-1421, 2019.

[8] B. Liang, W. Liu, L. Sun, Z. He, and B. Hou, "Economic MPC-based smart home scheduling with comprehensive load types, real-time tariffs, and intermittent DERs," IEEE Access, vol. 8, pp. 194373-194383, 2020.

[9] K. Amasyali, Y. Chen, and M. Olama, "A Data-Driven, Distributed Game-Theoretic Transactional Control Approach for Hierarchical Demand Response," IEEE Access, vol. 10, pp. 72279-72289, 2022.

[10] O. Erdinc, "Economic impacts of small-scale own generating and storage units, and electric vehicles under different demand response strategies for smart households," Applied Energy, vol. 126, pp. 142-150, 2014.

[11] O. Erdinc, N. G. Paterakis, I. N. Pappi, A. G. Bakirtzis, and J. P. Catalão, "A new perspective for sizing of distributed generation and energy storage for smart households under demand response," Applied Energy, vol. 143, pp. 26-37, 2015.

[12] H. Singabhattu, A. Jain, and T. Bhattacharjee, "Distributed energy resources optimization for demand response using MILP," in 2017 IEEE Region 10 Symposium (TENSYMP), July 2017, pp. 1-5.

[13] A. C. Duman, H. S. Erden, Ö. Gönül, and Ö. Güler, "A home energy management system with an integrated smart thermostat for demand response in smart grids," Sustainable Cities and Society, vol. 65, p. 102639, 2021.

[14] J. Aldahmashi and X. Ma, "Advanced Machine Learning Approach of Power Flow Optimization in Community Microgrid," in 2022 27th International Conference on Automation and Computing (ICAC), September 2022, pp. 1-6.

[15] M. Yousefi, A. Hajizadeh, M. N. Soltani, and B. Hredzak, "Predictive home energy management system with photovoltaic array, heat pump, and plug-in electric vehicle," IEEE Transactions on Industrial Informatics, vol. 17, no. 1, pp. 430-440, 2020.

[16] K. Garifi, K. Baker, B. Touri, and D. Christensen, "Stochastic model predictive control for demand response in a home energy management system," in 2018 IEEE Power & Energy Society General Meeting (PESGM), August 2018, pp. 1-5.

[17] F. Luo, G. Ranzi, C. Wan, Z. Xu, and Z. Y. Dong, "A multistage home energy management system with residential photovoltaic penetration," IEEE Transactions on Industrial Informatics, vol. 15, no. 1, pp. 116-126, 2018.

[18] H. Shuai, J. Fang, X. Ai, J. Wen, and H. He, "Optimal real-time operation strategy for microgrid: An ADP-based stochastic nonlinear optimization approach," IEEE Transactions on Sustainable Energy, vol. 10, no. 2, pp. 931-942, 2018.

[19] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," Applied Energy, vol. 235, pp. 1072-1089, 2019.

[20] O. Al-Ani and S. Das, "Reinforcement learning: theory and applications in hems," Energies, vol. 15, no. 17, p. 6392, 2022.

[21] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning," Energy, vol. 238, p. 121873, 2022.

[22] M. A. Khan, A. M. Saleh, M. Waseem, and I. A. Sajjad, "Artificial Intelligence Enabled Demand Response: Prospects and Challenges in Smart Grid Environment," IEEE Access, vol. 11, pp. 1477-1505, 2022.

[23] 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 Transactions on Smart Grid, vol. 14, no. 1, pp. 239-250, 2022.

[24] J. Lu, P. Mannion, and K. Mason, "A multi-objective multi-agent deep reinforcement learning approach to residential appliance scheduling," IET Smart Grid, vol. 5, no. 4, pp. 260-280, 2022.

[25] A. Mathew, A. Roy, and J. Mathew, "Intelligent residential energy management system using deep reinforcement learning," IEEE Systems Journal, vol. 14, no. 4, pp. 5362-5372, 2020.

[26] A. Forootani, M. Rastegar, and M. Jooshaki, "An advanced satisfaction-based home energy management system using deep reinforcement learning," IEEE Access, vol. 10, pp. 47896-47905, 2022.
[27] S. Sykiotis, C. Menos-Aikateriniadis, A. Doulamis, N. Doulamis, and P. S. Georgilakis, "Solar power driven EV charging optimization with deep reinforcement learning," in 2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), October 2022, pp. 1-6.
[28] J. H. Hong, D. Y. Hong, L. H. Yao, and L. C. Fu, "A demand side management with appliance controllability analysis in smart home," in 2020 International Conference on Smart Grids and Energy Systems (SGES), November 2020, pp. 556-561.
[29] A. Suleman, M. A. Amin, M. Fatima, B. Asad, M. Menghwar, and M. A. Hashmi, "Smart Scheduling of EVs Through Intelligent Home Energy Management Using Deep Reinforcement Learning," in 2022 17th International Conference on Emerging Technologies (ICET), November 2022, pp. 18-24.
[30] K. Kurte, K. Amasyali, J. Munk, and H. Zandi, "Comparative analysis of model-free and model-based HVAC control for residential demand response," in Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, November 2021, pp. 309-313.
[31] K. Kurte, J. Munk, O. Kotevska, K. Amasyali, R. Smith, E. McKee, Y. Du, B. Cui, T. Kuruganti, and H. Zandi, "Evaluating the adaptability of reinforcement learning based HVAC control for residential houses," Sustainability, vol. 12, no. 18, p. 7727, 2020.
[32] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," IEEE Internet of Things Journal, vol. 7, no. 4, pp. 2751-2762, 2019.
[33] Y. Yi, G. Verbič, and A. C. Chapman, "Optimal Energy Management Strategy for Smart Home with Electric Vehicle," in 2021 IEEE Madrid PowerTech, June 2021, pp. 1-6.
[34] T. Li, Y. Xiao, and L. Song, "Integrating future smart home operation platform with demand side management via deep reinforcement learning," IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 921-933, 2021.
[35] Y. Ye, D. Qiu, H. Wang, Y. Tang, and G. Strbac, "Real-Time Autonomous Residential Demand Response Management Based on Twin Delayed Deep Deterministic Policy Gradient Learning," Energies, vol. 14, p. 531, 2021.
[36] S. S. Shuvo and Y. Yilmaz, "Home energy recommendation system (hers): A deep reinforcement learning method based on residents' feedback and activity," IEEE Transactions on Smart Grid, vol. 13, no. 4, pp. 2812-2821, 2022.
[37] F. O. S. Saraiva and V. L. Paucar, "Locational Marginal Price Decomposition Using a Fully Distributed Slack Bus Model," IEEE Access, vol. 10, pp. 84913-84933, 2022.
[38] A. Kailas, V. Cecchi, and A. Mukherjee, "A survey of communications and networking technologies for energy management in buildings and home automation," Journal of Computer Networks and Communications, 2012.
[39] J. Hannagan, R. Woszczeiko, T. Langstaff, W. Shen, and J. Rodwell, "The Impact of Household Appliances and Devices: Consider Their Reactive Power and Power Factors," Sustainability, vol. 15, no. 1, p. 158, 2022.
[40] A. Ahmad, S. A. R. Kashif, M. A. Saqib, A. Ashraf, and U. T. Shami, "Tariff for reactive energy consumption in household appliances," Energy, vol. 186, p. 115818, 2019.
[41] S. Golshannavaz, "Cooperation of electric vehicle and energy storage in reactive power compensation: An optimal home energy management system considering PV presence," Sustainable Cities and Society, vol. 39, pp. 317-325, 2018.
[42] H. Arghavani and M. Peyravi, "Unbalanced current-based tariff," CIRED-Open Access Proceedings Journal, vol. 2017, no. 1, pp. 883-887, 2017.
[43] S. Zamanloo, H.A. Abyaneh, H. Nafisi, and M. Azizi, "Optimal two-level active and reactive energy management of residential appliances in smart homes," Sustainable Cities and Society, vol. 71, p. 102972, 2021.
[44] I. Mahdavi, B. Javadi, N. Sahebjamnia, and N. Mahdavi-Amiri, "A two-phase linear programming methodology for fuzzy multi-objective mixed-model assembly line problem," The International Journal of Advanced Manufacturing Technology, vol. 44, pp. 1010-1023, 2009.
[45] Y. F. Du, L. Jiang, C. Duan, Y. Z. Li, and J. S. Smith, "Energy consumption scheduling of HVAC considering weather forecast error through the distributionally robust approach," IEEE Transactions on Industrial Informatics, vol. 14, no. 3, pp. 846-857, 2017.
[46] E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, "Residential load and rooftop PV generation: An Australian distribution network dataset," International Journal of Sustainable Energy, vol. 36, no. 8, pp. 787-806, 2017.
[47] "World Weather Online," [Online]. Available: https://www.worldweatheronline.com. [Accessed: Jul. 20, 2022].
[48] Shengren, H., Vergara, P.P., Duque, E.M.S. and Palensky, P., "Optimal energy system scheduling using a constraint-aware reinforcement learning algorithm," International Journal of Electrical Power & Energy Systems, vol. 152, p. 109230, 2023.
[49] S. Uddin, H. Shareef, and A. Mohamed, "Power quality performance of energy-efficient low-wattage LED lamps," Measurement, vol. 46, no. 10, pp. 3783-3795, 2013.
[50] G. Gamrath, D. Anderson, K. Bestuzheva, W-K. Chen, L. Eifler, M. Gasse, et al., "The SCIP Optimization Suite 7.0," Optimization Online, Tech. Rep. 05, 2020.
[51] G. Wei, M. Chi, Z. W. Liu, M. Ge, C. Li, and X. Liu, "Deep Reinforcement Learning for Real-Time Energy Management in Smart Home," IEEE Systems Journal, 2023.
[52] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning," Energy, vol. 238, p. 121873, 2022.

Jamal Aldahmashi received the B.Sc. degree in Electrical Engineering from Northern Border University, KSA, in 2013, and the M.S. degree in Electrical Engineering from Denver University, USA, in 2018. He is currently pursuing the Ph.D. degree in Electrical Engineering with the School of Engineering, Lancaster University, U.K. His research interests include energy optimization, renewable energy sources integration, optimal power flow, and utilizing machine learning within smart grids.

Xiandong Ma received the B.Eng. degree in Electrical Engineering from Jiangsu University, China, in 1986, the M.Sc. degree in Power Systems and Automation from China State Grid Electric Power Research Institute in 1989, and the Ph.D. degree in Partial Discharge-based High Voltage Condition Monitoring from Glasgow Caledonian University, UK, in 2003. He is a Reader in Power and Energy Systems with the School of Engineering, Lancaster University, U.K.

His research interests include intelligent condition monitoring and fault diagnosis of the energy systems particularly with wind power generation, and modeling, optimization, and control of smart/micro grids with renewable energy resources. He is a Chartered Engineer (C.Eng.) and a Fellow of the Institution of Engineering and Technology (FIET).