

Electrical vehicle grid integration for demand response in distribution networks using reinforcement learning

Fayiz Alfaverh | Mouloud Denai  | Yichuang Sun

School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, UK

Correspondence

Mouloud Denai, School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, UK.

Email: m.denai@herts.ac.uk**Abstract**

Most utilities across the world already have demand response (DR) programs in place to incentivise consumers to reduce or shift their electricity consumption from peak periods to off-peak hours usually in response to financial incentives. With the increasing electrification of vehicles, emerging technologies such as vehicle-to-grid (V2G) and vehicle-to-home (V2H) have the potential to offer a broad range of benefits and services to achieve more effective management of electricity demand. In this way, electric vehicles (EV) become distributed energy storage resources and can conceivably, in conjunction with other electricity storage solutions, contribute to DR and provide additional capacity to the grid when needed. Here, an effective DR approach for V2G and V2H energy management using Reinforcement Learning (RL) is proposed. Q-learning, an RL strategy based on a reward mechanism, is used to make optimal decisions to charge or delay the charging of the EV battery pack and/or dispatch the stored electricity back to the grid without compromising the driving needs. Simulations are presented to demonstrate how the proposed DR strategy can effectively manage the charging/discharging schedule of the EV battery and how V2H and V2G can contribute to smooth the household load profile, minimise electricity bills and maximise revenue.

1 | INTRODUCTION

The ongoing challenge facing utilities across the world is to deliver electricity to customers during peak demand periods while keeping the power grid stable by balancing supply and demand of electricity. Today, most utilities have demand response programs (DR) in place to incentivise end-use customers to lower or shift their electricity usage at peak times. DR is essentially a class of Demand-Side Management (DSM), and has long been used by utility companies to optimise the operation of distribution grids while delivering reliable and cost-effective electricity supply to customers. Several DR approaches are adopted by utility companies and may be classified into two categories: incentives-based programs and price-based programs. In incentives-based programs, participating consumers are awarded fixed or time-varying payments for their consent to reduce energy consumption at peak hours or during contingency events [1].

Price-based DR is an indirect means for electricity suppliers to control consumers' electricity loads. In these programs, customers are charged with different rates based on the electricity price throughout different time periods. In this case, end-users have complete control over their loads and can adjust or schedule their demand in response to electricity price signals received from their energy supplier. Therefore, energy consumption can be reduced during peak hours when the prices are higher [2].

More recently, there has been a gradual shift towards the adoption of Electric Vehicles (EVs) in the automotive industry. The main drivers are the economic and environmental benefits, the technological improvement in batteries' energy density, government policies offering financial rewards such as tax breaks or rebates to EV owners [3]. In a report published by the UK National Grid in Future Energy Scenarios, the number of EVs is expected to grow significantly to 11 million by 2030 and to 36 million by 2040 [4]. EVs charging during peak hours leads to electricity price increase, additional demand and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Electrical Systems in Transportation* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

imposes severe stress on the distribution grid. This may create problems such as feeder congestion, distribution transformer overloads and excessive voltage drops which may impact the overall electricity network [5]. In contrast, off-peak EV charging benefits EV owners from lower electricity prices and helps reduce stress on the power grid. On the other hand, EVs could also serve as a temporary energy storage and supply power to home appliances during short-term outages, provide emergency charging to other EVs or feed power to the utility grid when needed. However, draining the complete EV battery energy during the day could potentially disrupt EV availability for travel needs [6]. Therefore, a more sophisticated management of the EV battery is required. As such, a holistic approach must be adopted for EVs to serve as a DR resource and close the energy gap. For example, in [7], a fair demand response with electric vehicles (F-DREVs) is proposed for a cloud-based energy management service to maximise incentives by minimising global cost within the given time period, and smooth fluctuations of EVs loads. In [8], a new scheduling approach is proposed for isolated microgrids (MGs) with renewable generations by incorporating demand response of EVs.

V2G technology has attracted a great deal of research interest in recent years both within the academic community and industry. A review on the impact of V2G technology is presented in [9]. In [10], different control schemes to enable EV grid integration are reviewed and the advantages and disadvantages of V2G integration are discussed with respect to the transient stability of the power grid at the transmission and distribution levels. The authors in [11] proposed a combined control and communication approach to ensure efficient energy transfer and maintain a balance between energy suppliers and consumers. Energy management of EV battery is proposed in [12] by taking into consideration the V2G connection, prices for selling and purchasing electricity, and the daily load profile of household appliances. In [13], an optimised smart charging and discharging coordination scheme using Linear Programming (LP) based on a heuristic algorithm is proposed for V2G technology. The aim of the proposed algorithm is to cope with a variety of loads and departure times of EVs. Bayesian Neural Networks (BNN) have been proposed in [14] for predicting electricity prices for charging/discharging EV battery while minimising charging costs over a long-term time horizon.

Classical optimisation methods such as linear programming, dynamic programming and their counterparts have been applied to the scheduling of EV charging/discharging. However, these methods suffer from the curse of dimensionality and cannot adapt to the environment's stochasticity including unpredictable load profile, price signals and changing driving patterns. Global search methods such as genetic algorithm, swarm intelligence and their hybrids with the linear optimisation methods are also used for solving power management problems. However, these methods are generally slow and computationally intensive; thus they are not suitable for real-time. In addition, when using such methods, which do not have a learning component, optimisation iterations are needed

for every new load and generation profile, which is also computationally intensive. Machine learning techniques such as reinforcement learning algorithms offer a better alternative as they can be trained offline for a general load and generation profile and then they can be applied online for any load profile, dynamic electricity price signal and various driving patterns.

Reinforcement Learning (RL) has been successfully used in various applications related to energy management, decision and control. RL models have excellent decision-making ability due to their ability to solve problems without a priori knowledge of the environment. Multi-agent reinforcement learning has been proposed for the optimal scheduling of household appliances including V2G technology in smart home to optimise the energy utilisation [15]. However, multi-agents RL requires setting several agents, with each agent having different actions and rewards therefore making the learning process more complex. Other studies have focussed on using Markov Decision Process (MDP) algorithms in HEMS to determine the optimal strategy for the scheduling of EV charging and discharging [16]. However, these algorithms require historical data, such as electricity prices and battery State-of-Charge (SOC) as inputs to compute the charging/discharging schedules in real-time.

Thus, the original contributions of the proposed method can be summarised as follows: (i) a new and flexible DR-based energy management strategy is proposed for EV charging and discharging operations without compromising the owner's driving needs and convenience. The proposed approach works with a single agent and uses a reduced number of state-action pairs and fuzzy logic for state space and reward functions instead of the classical rule-based techniques. (ii) the driving patterns are considered in this paper, where the EV model is interfaced to Google map via the App Designer Tool of MATLAB that enables the calculation of the distance, power required, and arrival and departure times for each trip. (iii) this study also quantifies the potential cost savings of various operating modes, including V2G, V2H, and G2V. (iv) this paper also evaluates the impacts of the participation of EVs in peak shaving within a residential area consisting of 100 homes. This model can be adjusted to any electricity network, independently of the country in which it may be located, and even if the demand conditions, electricity prices, and user's behaviour are different from the case study presented in this paper.

This paper is organised as follows: In Section 2, V2G and V2H technologies are briefly described. Section 3 presents the modelling of EVs. The concepts of RL and Q-learning model are presented in Section 4. Section 5 presents the results and discussion. Finally, the conclusions of the paper are summarised in Section 6.

2 | OVERVIEW OF V2G AND V2H TECHNOLOGIES

V2G and V2H are two technologies in which EVs having bidirectional power flow capability can connect to the charging station to draw power from the grid, deliver power to the grid

or provide back-up electricity supply to a home. The general energy flow diagram of charging/discharging modes of V2G and V2H structures within HEMS is illustrated in Figure 1.

A typical bidirectional EV battery charger consists of bidirectional AC/DC and DC/DC converters as shown in Figure 2 [17]. In charging mode, the bidirectional AC/DC converter is used to convert the AC grid power to DC for the battery and in discharging mode, the DC battery power is converted to AC power and injected back to the grid or used to supply the house. The DC/DC converter, on the other hand, controls the bidirectional power flow by using current control technique. The DC/DC converter can act as a buck or boost converter during charging or discharging mode, respectively.

With the advancement in battery technologies, the energy storage capacity of EVs has significantly improved. Currently, the capacity of EV batteries varies from 1 to 100 kWh. Battery capacity of Nissan Leaf 2018 is 40 kWh, that of Tesla Model 3 is 80 kWh and Tesla Model S is 100 kWh.

2.1 | Vehicle-to-grid (V2G)

V2G technology is attractive due to the benefits it provides to grid operators and EV owners and its positive impact on the environment. EVs with V2G capability are considered as an alternative energy source for the grid and can provide ancillary services to the grid such as frequency/voltage support, load balancing, support to intermittent power of renewable energy, reactive power support, valley filling and peak shaving. Therefore, an effective energy management is required to coordinate the charging/discharging modes of the EV battery. Smart chargers and their energy management system are key factors in the implementation of bidirectional V2G scheme. In [18], the authors proposed a controllable EV charger that enables an autonomous smart energy management system in a residential sector. The proposed power converter topology allows charging/discharging operations at different power levels. Several V2G demonstrator projects have been conducted around the world over the recent years most of them in Europe. Some of these V2G models have been adopted by leading car manufacturers and are already in the marketplace. In the UK, the Sciurus project is among the world's largest V2G projects aimed to develop and deploy a large number of chargers for domestic use. This project aims to validate the technical and commercial benefits of the V2G technology to the power grid and demonstrate its value to EV manufacturers [19]. In Germany, the world's leading car manufacturer Nissan, the transmission system operator TenneT and The Mobility House energy supplier have successfully completed a substantial V2G pilot project. The project aims to respond to the increasing concern of Germany about saving the surplus of energy generated from renewable, intermittent sources such as wind energy. In this project, Nissan Leaf batteries are used for energy storage. When fully charged, the batteries feed the stored energy

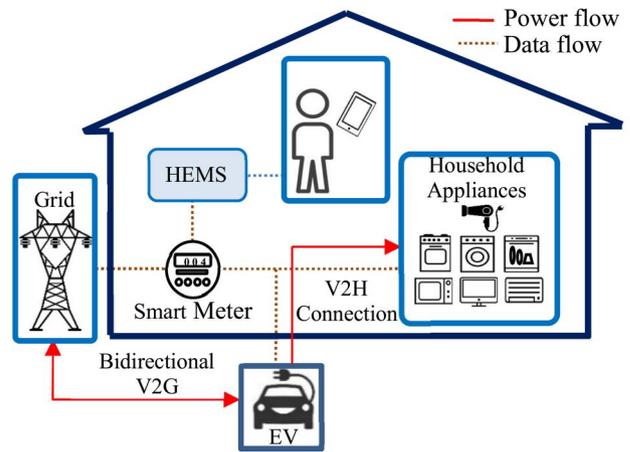


FIGURE 1 Block diagram of EV technologies integrated to HEMS. EV, electric vehicles

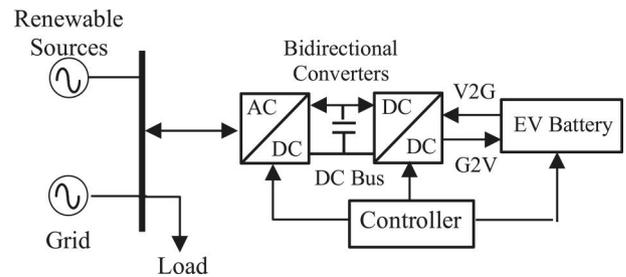


FIGURE 2 V2G using bidirectional converters. V2G, vehicle to grid

back to the grid when needed [20]. The SEEV4-City project, funded by the EU Interreg North Sea Region, aims to deploy V2G technology to use EV batteries as short-term storage of renewable energy to support the grid or redirect the available energy from vehicles to homes, neighbourhoods or cities [21].

2.2 | Vehicle-to-home (V2H)

V2H enables an EV to act as a backup power source and supply electricity to a home during short-term power outages or contribute to peak demand reduction, smooth home energy consumption, and minimise energy purchase from the grid. Some of the very significant and unique features of V2H technology is its simple implementation and some car manufacturers have already started deploying this technology. For example, Mitsubishi Motors Corporation (MMC) announced a new EV model called Dendo Drive House (DDH) at the 89th Geneva Motor Show [22]. This model is considered as a packaged system comprising an EV/PHEV, solar panel and a bidirectional V2H charger. DDH offers owners savings on charging costs and an emergency power source. Nissan Australia launched a new version of Nissan Leaf Plus that incorporates V2H capability [23]. V2H-equipped Leaf can be used as energy storage with the capability to supply energy to household appliances.

3 | MODELLING OF THE GRID INTEGRATION FRAMEWORK

3.1 | Modelling of V2G and V2H systems

Figure 3 depicts the overall structure of the EV model interfaced with the grid and the home and illustrates the operations modes and services it can provide during these modes.

Table 1 shows different types of EVs with different battery capacity and brands. The Nissan LEAF 2018 model is used in the simulations. The battery has a maximum capacity of 40 kWh which gives a driving autonomy of 151 miles, a charging time of 8 h to be fully charged (230 VAC 15 A).

The total power stored in the EV battery is given by the following equation:

$$E_{EV}^{Total} = E_{int} + \Delta E_{ch} - \Delta E_{dis-A} - \Delta E_{dis-grid} - E_{trip} \quad (1)$$

Where E_{EV}^{Total} is the net energy stored in the battery, E_{int} is the initial energy stored in the battery, ΔE_{ch} is the energy drawn from the grid to charge the EV battery, ΔE_{dis-A} represents the energy delivered to the household appliances, $\Delta E_{dis-grid}$ is the energy fed to the grid and E_{trip} denotes the total energy consumed by the EV during the trip.

$$\Delta E_{ch} = m_{G2V} \int_{t_{in}}^{t_{out}} P_{ch}(t) \cdot dt \quad (2)$$

$$\Delta E_{dis-A} = m_{V2H} \int_{t_{in}}^{t_{out}} P_{dis}(t) \cdot dt \quad (3)$$

$$\Delta E_{dis-grid} = m_{V2G} \int_{t_{in}}^{t_{out}} P_{dis}(t) \cdot dt \quad (4)$$

where $P_{ch}(t)$ and $P_{dis}(t)$ are the charging and discharging power, respectively. m_{V2H} , m_{V2G} and m_{G2V} represent the status of the EV connection mode and assume values 0 or 1.

The proposed management strategy depends also on the SOC of the battery which is the key parameter in the EV as it is a measure of the amount of the energy stored in it. The typical estimation of the SOC of the EV battery is based on the charging/discharging energy as follows:

$$SOC = SOC_{int} + \frac{\Delta E_{ch} + \Delta E_{dis-A} + \Delta E_{dis-grid}}{E_{EV}^{rat}} \quad (5)$$

where SOC_{int} is the initial SOC and E_{EV}^{rat} is the energy capacity of the battery.

Considering the battery lifecycle, some constraints have been imposed on the power delivered by the EV to the grid (V2G) or to the home (V2H) and on the battery State-of-

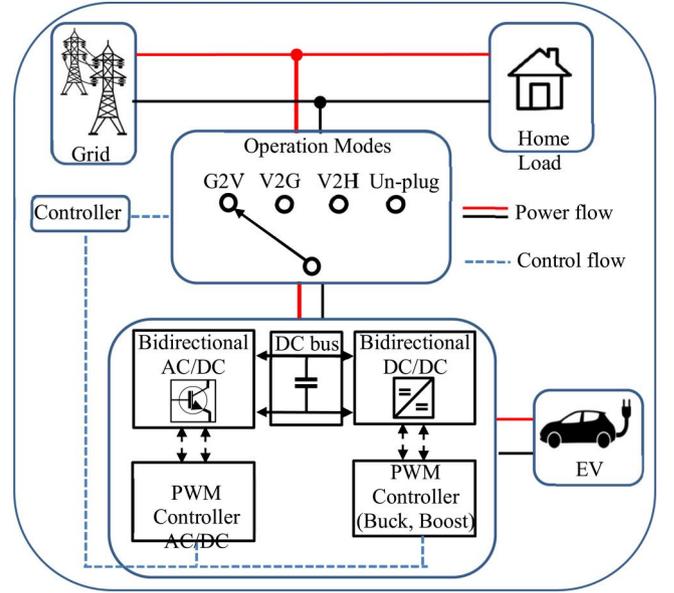


FIGURE 3 Selection of EV operation modes. EV, electric vehicle; G2V, grid to vehicle; V2G, vehicle to grid

Charge (SOC). Equation (6) shows that the EV power is limited between the minimum operating power P_{min} to be supplied to the EV and the maximum power range P_{max} to be injected in the grid/home.

$$P_{max} \geq P_{dis}, P_{ch} \geq P_{min} \quad (6)$$

Similarly, Equation (7) prevents deep discharging and full charging of the EV battery by imposing a minimum SoC (SOC_{min}) and a maximum SOC (SOC_{max}), respectively:

$$SOC_{min} < SOC < SOC_{max} \quad (7)$$

3.2 | Modelling of EV driving patterns

The EV model is interfaced to Google map via the App Designer Tool of MATLAB and allows the calculation of the distance, power required, arrival and departure time for each trip. This information will be employed for scheduling the battery charging and discharging times to ensure the vehicle is always sufficiently charged for the next trip.

The availability of EV refers to whether the vehicle is parked at home and accessible for either V2H or V2G connections and is defined as follows:

$$C_{ev} = \begin{cases} 1 & \text{EV available} \\ 0 & \text{EV not available} \end{cases} \quad (8)$$

TABLE 1 Different types of EVs

Manufacturer	Model	kWh	Miles	Battery Type
Ford	2018 Focus Electric	33.5	115	Liquid-cooled lithium-ion battery
Volkswagen	2018 Volkswagen e-Golf	35.8	120	Lithium-ion battery
Nissan	2018 Nissan LEAF	40	151	Lithium-ion battery
Chevrolet	2018 Chevy Bolt	60	238	Nickel-rich lithium-ion
Tesla	2018 Tesla model 3	75	285	Lithium-ion battery

Abbreviation: EV, electric vehicle.

Using the designed user-interface, the EV owner can schedule his/her trip by selecting the destination and departure times as shown in Figure 4.

Once the trip distance is determined, the energy required for the trip is calculated as follows:

$$E_{trip} = \frac{D_{trip}}{D_{max}} \times E_{EV}^{rat} \quad (9)$$

where D_{trip} and D_{max} are the distance of the trip and the maximum distance the EV can travel with full SOC, respectively, E_{EV}^{rat} denotes the maximum energy capacity of the EV battery.

3.3 | Modelling of the household load profile

Smart meters are used in smart home to receive the price signal from an energy supplier and collect the power data of all household appliances, and then send them to HEMS. Generally, the daily power consumption of a typical household contains two peaks occurring during the morning and evening times when energy prices are higher.

The off-peak period corresponds to the period of the day when electricity prices are lower since household activities such as washing, cleaning, cooking, and watching TV are reduced. In this work, the household appliances are divided into shiftable and non-shiftable appliances. Therefore, at each time step (considered as an hour in this work), the total power demand of all appliances during a certain hour is given as follows:

$$E_b^T = \sum_{n=1}^N e_n^{shft} \cdot J_n^{shft} + \sum_{m=1}^M e_m^{non} \cdot J_m^{non} \quad (10)$$

Where e_n^{shft} and e_m^{non} represent the rated power of each shiftable appliance and non-shiftable appliance, respectively. $J_{n/m}^{shft/non}$ denotes the status of the appliance and takes values 0 (off) or 1 (on), respectively, $t \in \{1, 2, 3, \dots, 24\}$ represents the hour of the day, $n \in \{1, 2, \dots, N\}$ is the appliance number and N is the total number of the shiftable appliances. $m \in \{1, 2, \dots, M\}$ refers to the appliance number and M is the total number of the non-shiftable appliances.

4 | V2G DEMAND RESPONSE BASED ON Q-LEARNING

With V2G technology, EVs become a distributed energy storage and can offer a range of ancillary services by back-feeding power to the electricity grid such as demand response, peak-load management, voltage support, frequency regulation and will enable EV owners to save on energy costs and generate revenue through price arbitrage. V2G technology is evolving at an accelerated pace and research is ongoing to enhance its functionality and implementation.

This paper proposes an effective V2G-based demand response approach for the energy management of residential loads using Q-learning and RTP.

4.1 | Overview of Q-Learning

Reinforcement Learning (RL) is a machine-learning type of computational algorithm in which an agent tries to maximise the total reward over time by taking actions and interacting with an unknown environment.

RL algorithm works based on main six parameters namely, agent, environment, state space \mathcal{S} , action space \mathcal{A} , rewards \mathcal{R} , and action-value $Q(s, a)$. Generally, the RL-agent interacts with an environment as shown in Figure 5.

At each time step $t = \{0, 1, 2, \dots\}$ the agent takes an action $a_t \in \mathcal{A}(t)$ based on a certain policy π at a given state $s_t \in \mathcal{S}(t)$. The environment then observes the new state $s_{t+1} \in \mathcal{S}(t)$ and computes a numerical reward $r(s_t, a_t)$ to evaluate the performance of the action taken which is then fed back to the agent as shown in Figure 5. Based on the calculated reward, the agent will optimise its policy π and hence maximise the accumulated future rewards.

In RL algorithm, the action-value function refers to ‘‘how good’’ the action taken performs for a certain state based on a certain policy π , and is denoted as $Q_\pi(s, a)$. This action-value function is defined as the mathematical expectation of the total rewards that the agent will obtain in the future:

$$Q_\pi(s, a) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right] \quad (11)$$

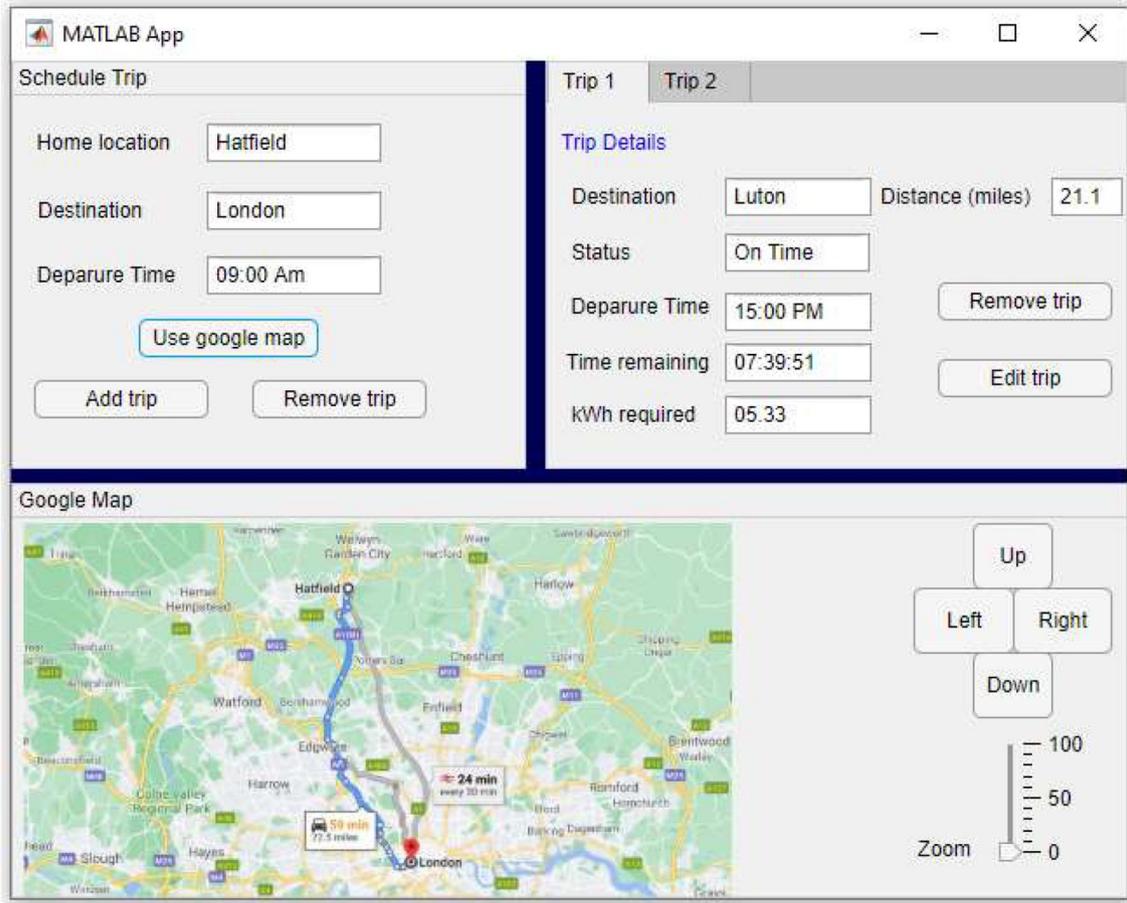


FIGURE 4 User-interface for scheduling trips

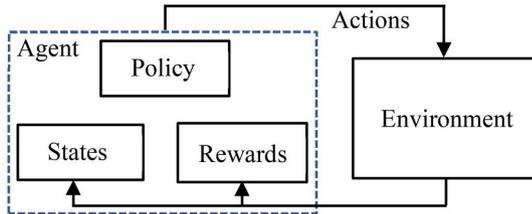


FIGURE 5 Reinforcement learning process

Where E_{π} denotes the expectation of total rewards by following policy π , γ is known as the discount rate and describes the relationship between the current and future rewards. It takes a value in the interval $[0, 1]$, where '0' means the agent relies on the current reward only and '1' indicates that the agent will strive for the future rewards.

At each state, the agent can take at least one optimal action and receive the highest reward. Therefore, the optimal policy aims to choose the optimal action with the highest Q-value as follows:

$$\pi(a|s) = \operatorname{argmax}(Q(s, a)) \quad (12)$$

Q-learning is a model-free RL algorithm that aims to learn a policy which guides the agent towards the optimal action

under certain states. The procedure of Q-Learning is to create a Q-matrix that has a dimension of $S \times A$ and then assign a Q-value $Q(s_t, a_t)$ to each state-action pair at time step t , and then update this value at each iteration to optimise the agent's performance.

The action-value $Q(s_t, a_t)$ is updated using the following equation:

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha[r(s_t, a_t) + \gamma \cdot \max_{a'} Q(s_{t+1}, a_{t+1})] \quad (13)$$

where α ($0 < \alpha < 1$) denotes the learning rate and determines how much the new reward affects the old value of $Q(s_t, a_t)$. For instance, $\alpha = 0$ means that the new information obtained is ignored and hence the reward received does not affect the Q-value. When $\alpha = 1$, only the latest information is considered.

4.2 | Q-learning model for the EV

RL is adopted here to make an optimal decision on charging or discharging of EV battery under dynamic electricity prices and different energy consumption patterns using an intelligent

agent which controls the dynamic process by executing sequential actions. The dynamic process is characterised by a state-space and a numerical reward that evaluates the new state when a given action is taken. In this paper, the Q-learning model components are defined as follows:

4.2.1 | State-space implementation using fuzzy logic

To reduce the number of states, fuzzy logic toolbox of MATLAB software is used. The state-space here is represented by the total household loads profile, SOC of the EV battery, availability of the EV and the electricity price signal. To simplify the model and reduce the computation time, the household power demand is divided into four levels as *Extremely Low*, *Low*, *High* and *Extremely High*. The SOC of the EV battery is defined as *Extremely Empty*, *Low*, *High* and *Extremely Full*; the EV availability can be *Available* or *Unavailable*. Finally, the price signal is categorised into *Cheap* and *Expensive* price.

The state is formulated using Fuzzy Logic. Fuzzy reasoning is a decision-making model that deals with approximate values rather than exact values. A Fuzzy Inference System (FIS) provides the mapping from the inputs to the outputs, based on a set of fuzzy rules and associated fuzzy Membership Functions (MFs). There are two types of FIS, Mamdani-type FIS and Sugeno-type FIS. Mamdani method is used in this paper because it offers a smoother output. The input variables of the fuzzy state model are the total home power demand (E_t^{total}), the electricity price (P_t), the SOC of the EV battery (SOC_t) and the availability of EV (C_v), and the output variable is the States.

The MFs for the input variable ‘Total household loads’ are triangular and are labelled as *Extremely Low*, *Low*, *High* and *Extremely High*. The universe of discourse of power demand is chosen as [0 6300] (Watt). The fuzzy sets of electricity price are defined as *Cheap* and *Expensive*. The MFs are Gaussian and the universe discourse is [0 0.16] (£/kWh). For the SOC, the universe discourse is defined [0 100]%, and the associated fuzzy sets are *Extremely Empty*, *Low*, *High* and *Extremely Full*. The output of the system is the State, and there are 25 States in total.

4.2.2 | Action space

The proposed strategy aims to schedule the charging time of the EV to be during off-peak hours when electricity prices are lower. It also aims to manage the energy stored in the EV battery whether to supply the household appliances during high energy prices or outage (V2H), or to deliver the energy back to grid using (V2G) depending on the SOC and availability of the EV.

Therefore, the actions set can be summarised as Charging, Discharging/Appliances, Discharging/Grid and Do nothing. Based on the current state, the agent chooses one action from the action space A that given by -

$$A = \left[\text{Charging}, \text{Discharging}_{\text{Appliances}}, \text{Discharging}_{\text{Grid}}, \text{Donothing} \right] \quad (14)$$

4.2.3 | Reward function implementation using fuzzy logic

The agent receives a numerical reward $r(s_t, a_t)$ after executing a random action and observing the new state. This reward value aims to evaluate how good is the action taken by the agent for a certain state. Fuzzy logic is also used here to evaluate the action that will be taken for a certain state.

The input variable of the reward function's FIS is the current state s_t , which is the output of the FIS system of the state space. The outputs of the system are the evaluation of the random action which was defined in Q-learning. For each action taken (output), the fuzzy sets are determined as Bad Action (BA), Good Action (GA) and Very Good Action (VGA). The universe of discourse of MFs is defined as [0 100] to evaluate all possible actions with values out of 100. The evaluation actions process works as follows; first, the FIS system of the state-space will identify the current state, based on the current space, the agent will take a random action from the action space. The FIS system of the reward function evaluates all the possible actions with value out of 100. Then, the agent will receive a numerical value corresponding to the action taken.

4.3 | EV energy management strategy using Q-learning

Q-learning is an off-policy RL algorithm that aims to make the appropriate decision at a current state. Using the off policy, the agent learns from taking a random action at a certain state without following a current policy. This means that a policy is not required during a training process. The Q-matrix with a dimension of [states \times actions] should be initialised to zero (i.e. the Q-value of each state-action pair is signed to zero). Then, the agent will interact with the environment and update each pair in that matrix after each action taken using Equation (3). In this paper, a random action called ‘exploring’ is applied. In this case, an appropriate number of iterations will be required to explore and update the values of $Q(s_t, a_t)$ for all state-action pairs at least once. After convergence of the Q-matrix, the optimal Q-values will be obtained.

The pseudo-code listed in Table 2 (Algorithm 1) illustrates the procedure of the main algorithm of the EV energy management using Q-learning based on MATALB/Script. First, a certain state and a numerical reward are defined using fuzzy logic. The parameters γ and α are set to 0.8 and 0.2, respectively, and Q-value matrix entries are initialised to zeros. For each current state, all possible actions are specified, and then an action will be selected randomly. After the selected action is executed, the numerical reward (using fuzzy logic) for that action and the new state will be observed by the agent. The maximum Q-value for the next state should also be determined

and then the Q-value of the state-action pair will be updated using Equation (13).

Finally, the next state will be used as a current state. To allow the agent to visit all state-state pairs and learn new knowledge, the training process is set to 1000 iterations.

5 | RESULTS AND DISCUSSION

5.1 | Implementing the proposed strategy with single EV in a household

The proposed EV management strategy works based on the relationships between the household power demand, electricity price, SOC and availability of EV. The smart meter is used in the smart house to collect the power data of the household loads, receive the energy price signal from the power grid and send them to HEMS. Once the EV arrives home, HEMS will receive a signal of EV availability with the percentage of SOC as illustrated in Figure 6. Consequently, HEMS can make an optimal decision whether to charge the EV battery from the grid, use it to power the household

appliances or sell the energy stored in the battery back to the grid. The simulation time is set at one day (24 h) with a 5-min time step. The price tariff signal that is received from the power utility is shown in Figure 7. It is assumed that the user leaves the home by 09:00 and returns at 14:00.

The total power demand of the household appliances is shown in Figure 8. There are two peak periods, morning peak hours [7:00 AM–10:00 AM] when the household member wake up and evening peak period [18:00 PM–21:00 PM] when the users start cooking, watching TV and other activities. The mid-peak period occurs after and before these two peak periods. Off-peak period occurs usually after midnight until morning.

5.1.1 | Case 1: weekdays with 70% of SOC

In this case, the initial SOC of the EV battery is set to 70%, based on the relationship between the electricity price signal, the total power demand and the Q-values. Figure 9a show all actions taken. For example, the total energy demand is low (3.5 kWh) during the period [5:00–7:00], the energy price is 0.07 £/kWh at 5:00 AM and 0.08 £/kWh at 6:00 AM, the SOC is high at 70% and the EV is available $C_{EV} = 1$. Therefore, the current state is defined as $s_t = [Low, Cheap, High, 1]$. According to the Q-value, the maximum Q-value for this state refers to the Charging action. To protect the battery from the overcharging, the maximum SOC is set to 80%. Therefore, from Figure 9b, it can be observed that the charging mode is stopped during the last half an hour because the SOC reached 80%, and the Do-Nothing mode is active. At 7:00 AM, the system moves to another State because the energy demand is average (4.3 kW), the price is high (0.11) and the battery is fully charged with SOC of 80%. Therefore, it is better to sell energy to the grid using V2G connection during this hour. In the next hour (8:00 AM), the demand jumps to 4.8 kW (High), the energy price is 0.14 £/kWh and the SOC remains high at 70%, then the best decision is to supply the household appliances using V2H connection during this hour.

The user leaves his house with the EV at 9:00 AM and returns at 14:00 PM, during this period the EV is not available, hence Do-Nothing mode is issued. When the EV returns home at 14:00 PM, the SOC is low, and the energy demand and price

TABLE 2 Algorithm of the EV energy management using Q-learning

Algorithm 1	
- Set γ, α parameters and environment rewards.	
- Initialise $Q(s_t, a_t)$, $\forall s \in S, \forall a \in A$.	
For each time step t do:	
- Choose a random initial state.	
While hour = 1:24	
- Determine all available actions.	
- Select random action from all possible actions for the current state.	
- Execute the selected action a_t , and observe the new state s_{t+1} and numerical reward $r(s_t, a_t)$.	
- Determine the maximum Q-value for next state in Q-matrix.	
- Update the $Q(s_t, a_t)$ using Equation (13).	
- Set the next state as current state.	
End while	
End for	

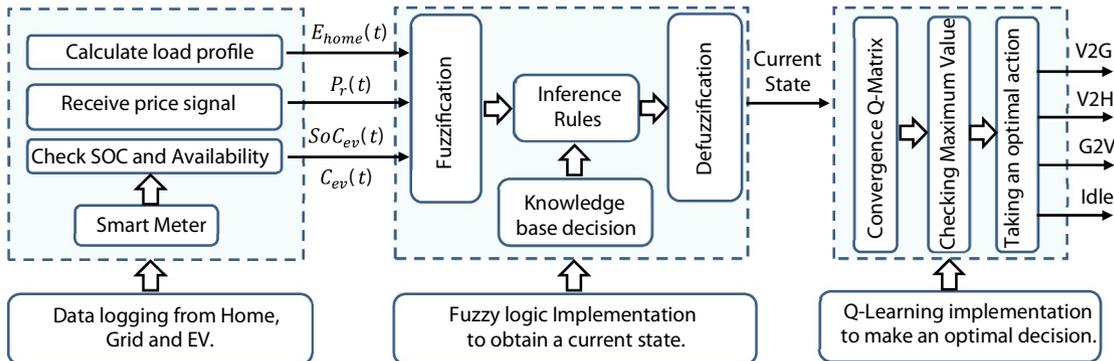


FIGURE 6 Flowchart describing the proposed method. G2V, grid to vehicle; V2G, vehicle to grid

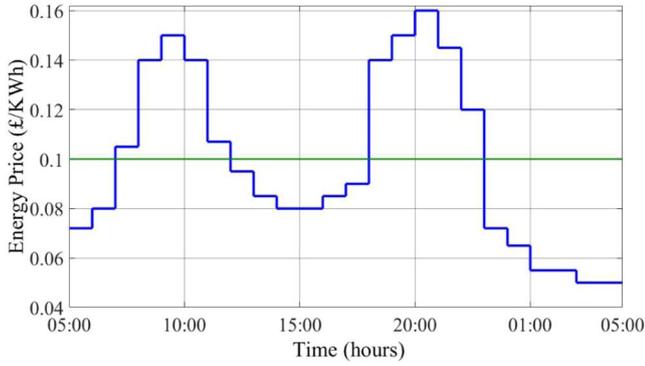


FIGURE 7 Real-time price (blue) and average price (green) signals

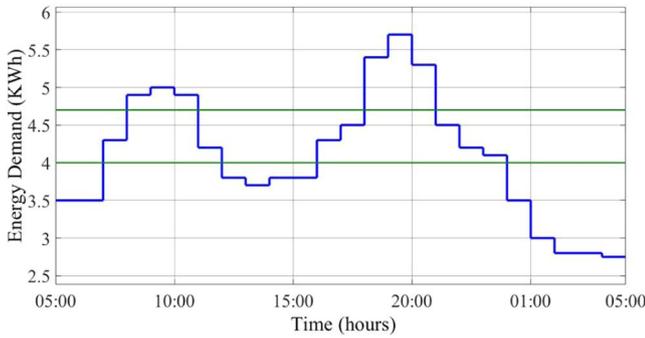


FIGURE 8 Total power demand of the smart home

are also low (3.8 kW and 0.08 £/kWh, respectively). Therefore, the EV battery is charged for two hours from 14:00 PM–16:00 PM till the SOC reached 60%. During the next two hours (16:00–18:00 PM), the action taken is Do-Nothing because the SOC is high, the energy price is low, and the energy demand is average. Therefore, there is no need to charge or discharge the battery. During the evening peak period from 18:00 PM to 21:00 PM, the energy demand and price are quite high, the EV battery has 60% of SOC. Discharging/Appliances mode is triggered to supply energy to the appliances and reduce the electricity cost as shown in Figure 9c. Finally, during the period from 21:00 PM to 23:00 PM, Discharging/Selling mode is activated to sell energy back to the grid since the SOC is at 50%. This period leads to maximise the user's revenue and reduce the stress on the power utility. The SOC drops to 30% at 24:00, hence Do-Nothing mode occurs. The Charging mode is activated at 24:00 PM till 5:00 AM, and the SOC reaches 60% to be ready for the next day.

5.1.2 | Case 2: weekdays with 50% of SOC

In this case, the initial SOC of the EV battery is set to 50%, because the SOC is reduced by 20% from case 1, the amount of energy sold to the grid is also reduced. Figure 10a shows all the actions taken. The EV starts the day with a SOC of 50% as shown in Figure 10b. During the first two hours [5:00 AM–7:00 AM], the Charging mode is active until the SOC reaches

60%. During the morning peak hours [7:00 AM–9:00 AM], the EV battery is used to inject power to the grid at 7:00 AM and supply the household appliances with different states and actions at 9:00 AM as shown in Figure 10c. When the EV returns home at 14:00 PM, the mode Charging is activated for hours [14:00 PM–16:00 PM] to charge the EV battery with lower price and to have it available for the evening peak period. The V2H model is active during [18:00 PM–21:00 PM] to deliver energy to the household appliances using Discharging/Appliances action. From 24:00 PM to 5:00 AM (off-peak period), the EV starts charging to be ready for travelling purposes in the next day, the SOC reaches 60%.

5.1.3 | Case 3: weekdays with 30% of SOC

The initial SOC of the EV is 30% in this case. All actions taken during this scenario are shown in Figure 11a. The charging action is active during [5:00 AM–7:00 AM] as shown in Figure 11b till the SOC reaches 40%. Although the SOC is lower, the battery is used as much as possible to supply the home appliances and sell power to the grid. This minimises the electricity bills and reduces the burden on the grid. Therefore, the EV exports energy to the grid only for one hour [7:00 AM–8:00 AM] in this case, The EV also supplies energy to the appliances for one hour during the morning peak period [8:00 AM–9:00 AM] as shown in Figure 11c. After midnight, the EV starts charging and the SOC reaches 50%.

5.1.4 | Case 4: Weekend days

Figure 12 illustrates the hourly power demand during weekends and weekdays. Hourly power consumption patterns of weekdays and weekends are quite different. The weekend pattern has a clear morning peak start at 10:00 AM and lags weekday pattern. The different features are quite understandable because people usually start later during weekend mornings and generally spend more hours at home during the time slot from 10:00 AM to 14:00 PM while they spend more time at work or on outdoor activities at the same time slot during weekdays. In this scenario, it is assumed that the EV leaves the home twice [13:00–15:00] and [20:00–22:00], the EV starts the day with 50% of SOC.

During the time interval [5:00 AM–7:00 AM], Charging mode is detected as shown in Figure 13a as the power demand (3 kWh) and electricity price are lower. The SOC increases from 50% to 70% as shown in Figure 13b. Because it is the weekend day, EV owner is staying at home, the SOC is quite high and power demand is low, it is better to sell energy to grid. Therefore, Discharging/Selling mode is active from 7:00 AM–10:00 AM and the SOC decreases to 60%. The next three hours [10:00 AM–13:00 PM], the power demand is high as it is morning peak demand. Discharging/Appliances mode is detected to supply power to the home appliances. The EV leaves the home at 13:00 PM and comes back at 15:00 PM with low SOC (38%). Hence, Charging mode is detected for three

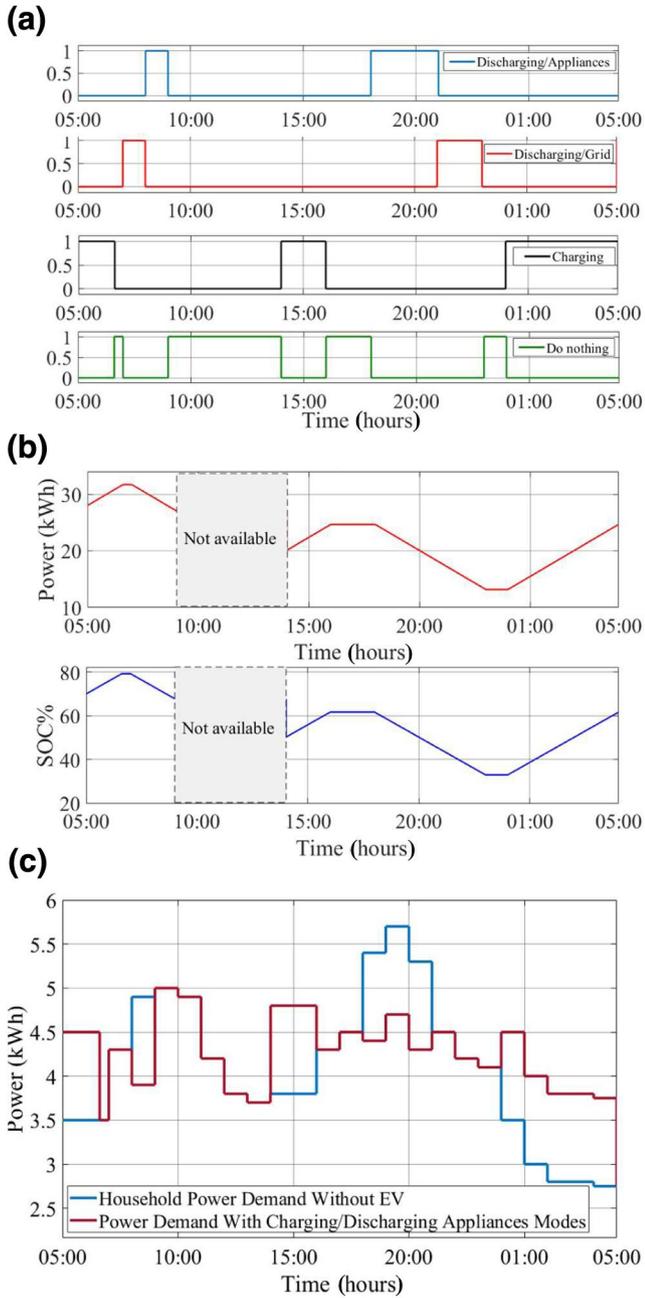


FIGURE 9 Initial SOC = 70% (a) Actions, (b) SOC changes, (c) Power consumption. SOC, state of charge

hours [15:00 AM–18:00 PM] and the SOC increases to 65%. During [18:00 PM–20:00 PM], the SOC is quite high, and the electricity price is high, the demand is average, hence Discharging/Selling mode is active to feed energy back to the grid. The EV leaves the home at 20:00 PM and come back at 22:00 PM. In that time, the evening peak demand begins.

Therefore, the EV starts supply the household appliances for two hours till SOC drops to 41%. At 24:00 PM, the off-peak demand is detected with low energy price. The EV starts charging until it reaches 80% at 04:00 AM and then Do-Nothing mode is active.

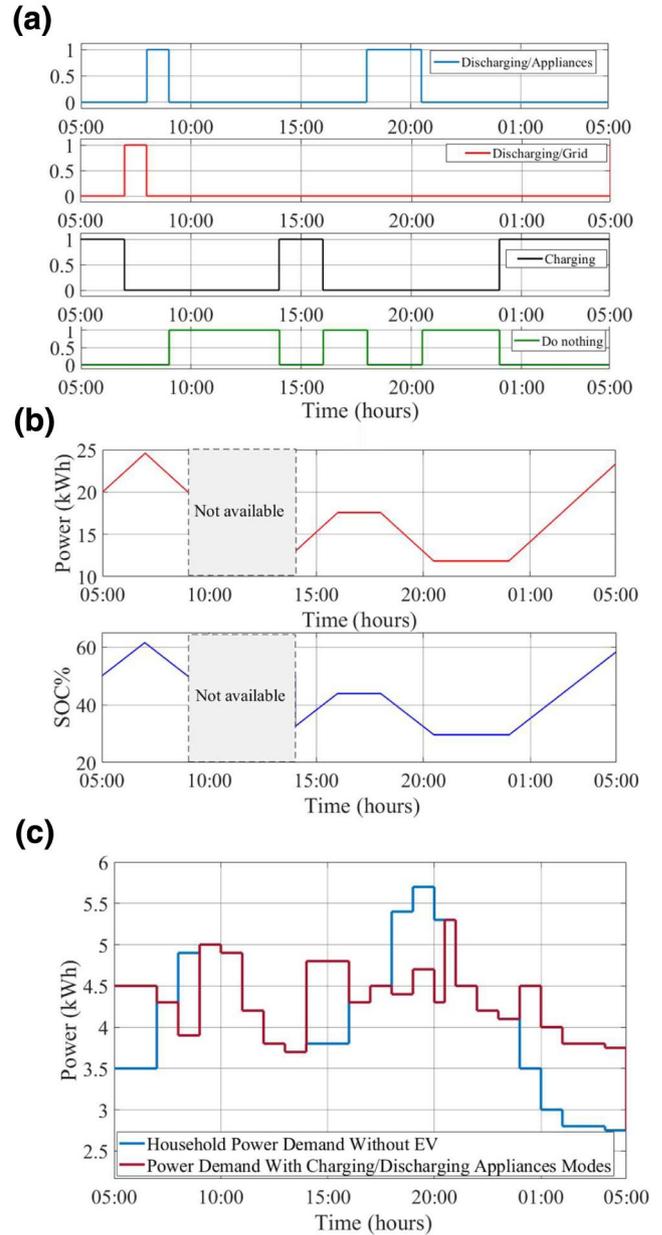


FIGURE 10 Initial SOC = 50% (a) Actions, (b) SOC changes, (c) Power consumption. SOC, state of charge

Figure 14 shows the cumulative daily household energy cost for the base case and different cases (SOC = 30, 50% and 80%). It can be observed from the figure that the proposed method reduces the cost by 12% when the EV starts the day with low SOC (30%), 21% of the cost reduction when the SOC is 50%. When SOC is higher (80%), the energy cost reduces by 27%.

5.2 | Implementing the proposed strategy on a fleet of EVs connected to the distribution network

A fleet of EVs with bidirectional power capability represents a considerable energy storage resource and has the potential to

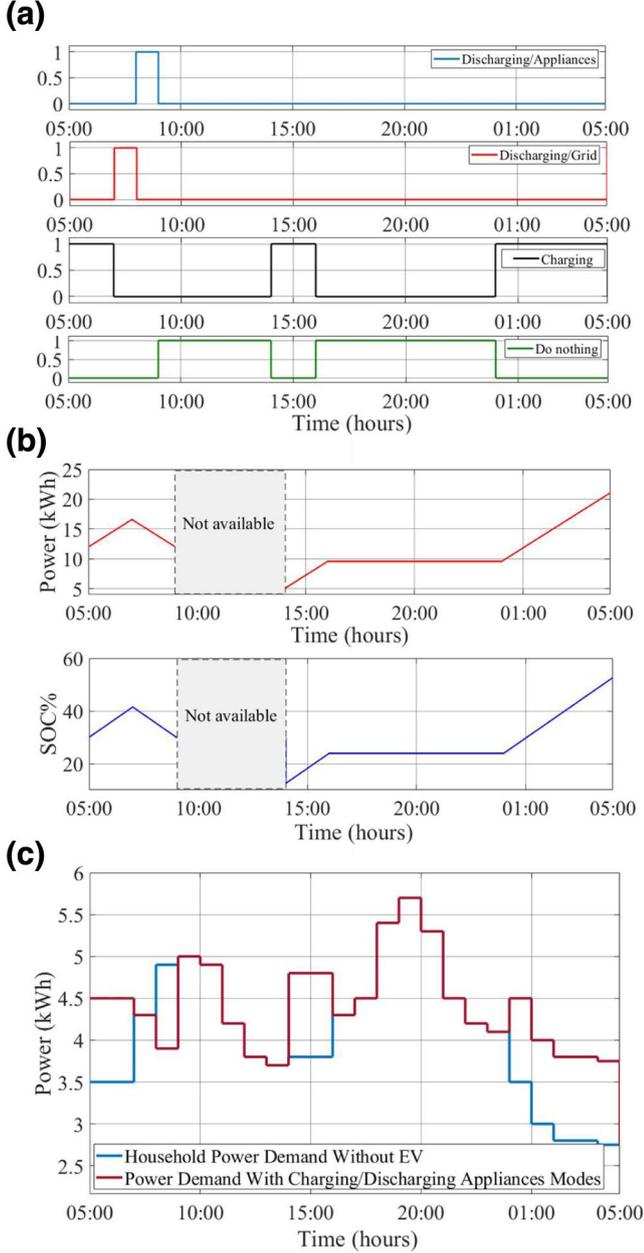


FIGURE 11 Initial SOC = 30% (a) Actions, (b) SOC changes, (c) Power consumption. EV, electric vehicle; SOC, state of charge

provide ancillary services to the utility grid. To evaluate the effectiveness of the proposed scheduling scheme, the low-voltage (LV) distribution network connected to a residential area is used. The residential area consists of 100 houses with 50 households owning an individual EV. There are different EVs with different battery capacity in the market, therefore different EVs are selected randomly among [20, 30, 40 and 50 kWh] as presented in Figure 15a. The random initial SOC for each EV is chosen. Figure 15b shows the number of EVs with different SOC.

In this model, three variables are defined for each EV which are the departure time, arrival time and energy consumption during a trip. In addition, for simplicity, it is assumed

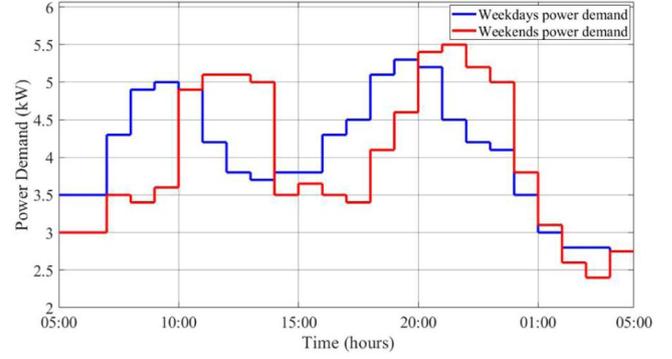


FIGURE 12 Hourly household electrical demand of weekend and weekday

that each EV performs only one trip during a day. However, the model can be easily extended to accommodate multiple trips for each EV. These three variables are modelled with a probability density function having a normal distribution. The intervals for departure and arrival times are [7 AM–10 AM] and [16 PM–22 PM], respectively, as shown in Figure 15c. The energy consumed by each EV during a trip is between [0–10 kWh].

The energy consumption in the residential area in real time is presented as follows:

$$E_{MG}(t) = E_{bs}(t) + \Delta E_{cb}^{N_{cb}}(t) - \Delta E_{dis}^{N_{dis}}(t) \quad (15)$$

where $E_{MG}(t)$ is the consumed energy in the residential area. $E_{bs}(t)$ refers to the energy consumption by households' appliances. $\Delta E_{cb}^{N_{cb}}(t)$ and $\Delta E_{dis}^{N_{dis}}(t)$ are the charging and discharging energy by EVs, respectively.

$$E_{bs}(t) = \sum_{n=1}^{N_b} E_b^T \quad (16)$$

$$\Delta E_{cb}^{N_{cb}}(t) = \sum_{n=1}^{N_{cb}} P_{cb}(t) \quad (17)$$

$$\Delta E_{dis}^{N_{dis}}(t) = \sum_{n=1}^{N_{dis}} P_{dis}(t) \quad (18)$$

Figure 16 shows the power consumption profile of the residential area in a typical weekday without EVs charging load. The peak hours occur during morning [7:00 AM–11:00 AM] and evening [18:00 PM–21:00 PM].

Figure 17 shows the power profile curve of the residential area after implementing the proposed strategy for each EV in the network. Some of the users start charging their EVs during off peak [12 AM–7 AM] based on the SOC, departure time of each EV, this results in filling the valley. It can be observed that

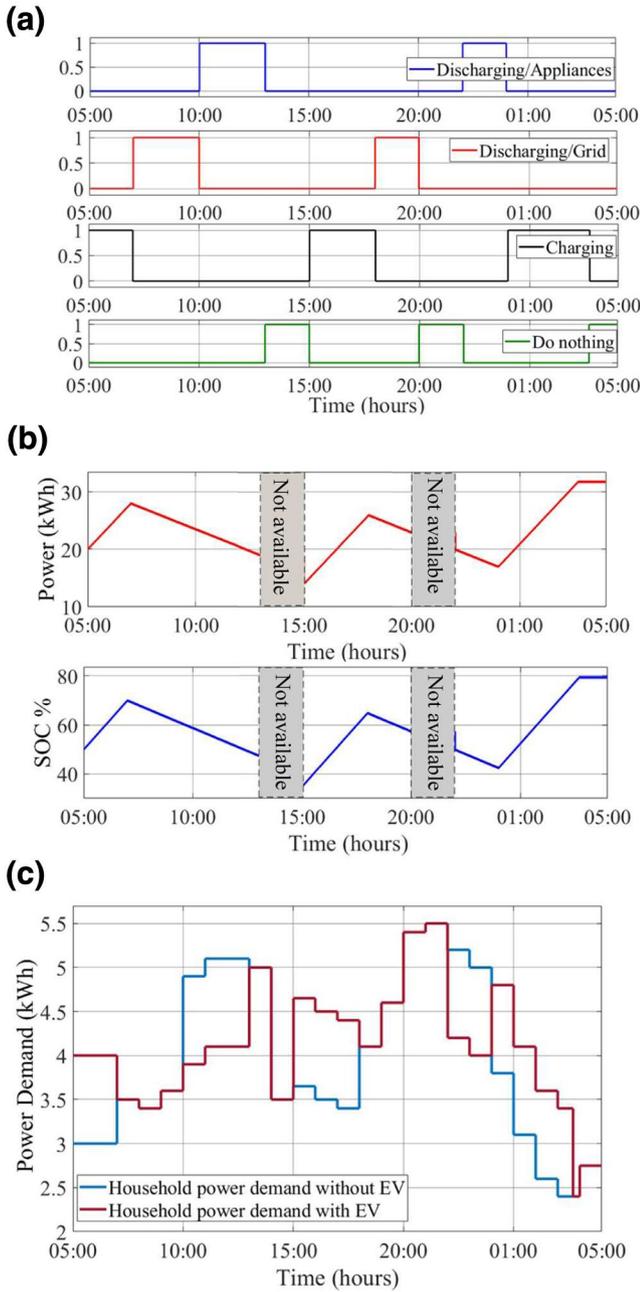


FIGURE 13 Weekend days with Initial SOC = 50% (a) Actions, (b) SOC changes, (c) Power consumption. SOC, state of charge

the morning peak and evening peak periods have been reduced by 23% and 15%, respectively. This is the result of EVs feeding power back to the grid via V2G and supplying energy to the household to power the appliances via V2H which reduced the energy consumed from the grid.

Table 3 summarises the results of this simulation. The total peak reduction during morning hours is 23% (15% using V2H and 8% using V2G) and a reduction of 15% occurred at evening (10.49% using V2H and 4.51% using V2G). It can be observed that EVs with higher initial SOC have the highest probability to feed energy back to the grid and supply the household appliances. Therefore, the number of EVs that start

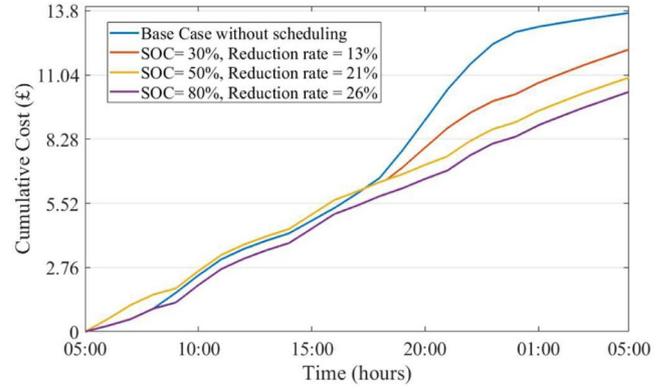


FIGURE 14 Cumulative electricity cost. SOC, State of charge

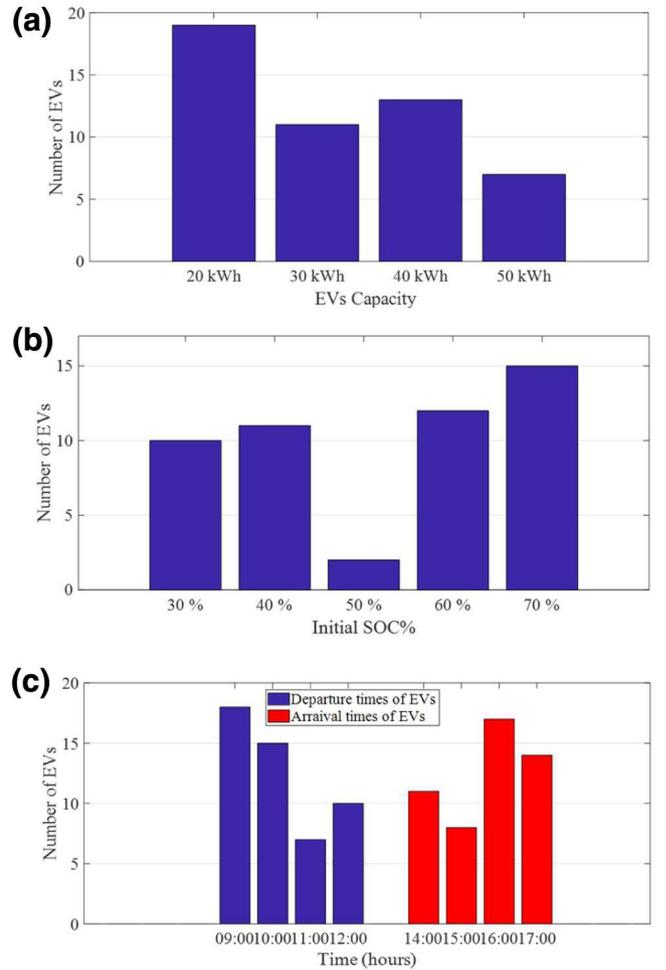


FIGURE 15 Number of EVs with different (a) capacity (b) initial SoC (c) departure and arrival times. EV, electric vehicle; SOC, state of charge

the day with 70% is 15. These EVs have the highest contribution in reducing the morning peak period by 7.20% and 6.0% during evening peak hours. However, the EVs with lowest of initial SOC (30%) could not sell energy during both peak periods (morning and evening), but these are still able to supply energy to the household appliances and contribute to load reduction (1.61% at morning and 1.05% at evening).

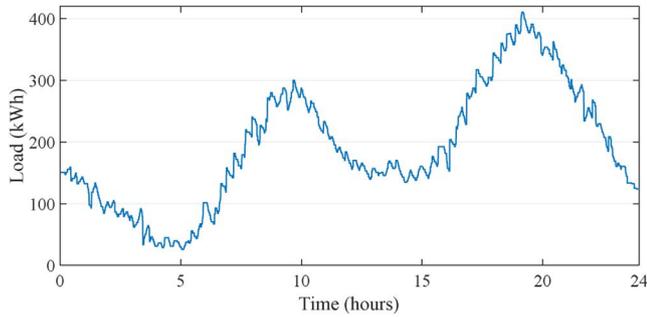


FIGURE 16 Load profile of the residential area. V2G, vehicle to grid; V2H, vehicle to home

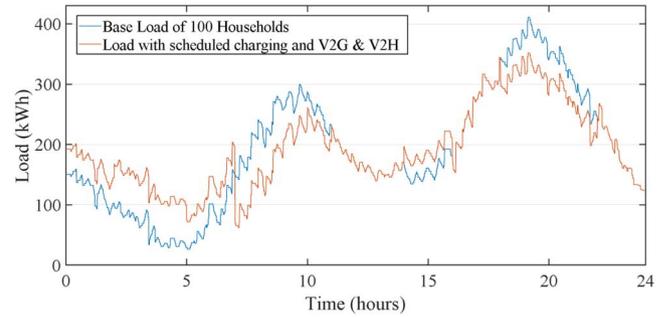


FIGURE 17 Network power profile after implementing the proposed strategy

TABLE 3 Contribution of EVs to energy reduction during morning and evening peak periods

SOC	No. of EVs	EVs Capacity (kWh)				Reduction of peak loads			
		20	30	40	50	Morning		Evening	
						By V2H	By V2G	By V2H	By V2G
70%	15	7	1	5	2	5.20%	4.00%	3.90%	2.10%
60%	12	2	6	1	3	4.5%	2.40%	2.35%	2.15%
50%	2	0	1	1	0	0.59%	0.10%	0.28%	0.17%
40%	11	6	1	2	2	3.10%	1.50%	2.91%	0.09%
30%	10	4	2	4	0	1.61%	0.00%	1.05%	0.00%
Total	50	19	11	13	7	15.00%	8.00%	10.49%	4.51%
						23%		15%	

Note: The bold is only to emphasize the totals.

Abbreviations: EV, electric vehicles; V2G, Vehicle to grid; V2H, vehicle to home.

These simulations have been performed using a laptop computer with an Intel(R) Core (TM) i5-3320M CPU @ 2.60 GHz and 16 GB of RAM. The average computational time required to determine the optimal policy of the Q-learning is 250 milliseconds with 10,000 iterations.

6 | CONCLUSION

This paper proposed a demand response strategy for charging/discharging energy management of Electrical Vehicles (EV) equipped with bidirectional V2G and V2H functionalities. The proposed strategy aims to minimise the household energy consumption and EV charging costs by charging the battery during off-peak hours when energy prices are lower or discharging (supplying) the home appliances during peak periods or support the electricity grid during peak demands which also generates revenue to the EV owner. In this study, an effective EV energy management system is developed using Q-learning to coordinate the charging/discharging modes considering the travel needs of the EV owner. The Reinforcement Learning (RL)-based scheme is employed, where the EV is defined as an agent and utilises a signal agent to make an optimal decision. A signal agent approach is used to reduce the number of state-action pairs, which simplifies the implementation, leads to a

better performance, and lower time consumption as compared with other techniques. Fuzzy reasoning is introduced to define the current state based on the input variables and evaluate the random action that the agent could take as a reward function. Fuzzy logic-based implementation of the state and reward function overcomes the limitation of rules-based technique (crisp values) and leads to a better performance. The proposed strategy was also successfully applied to a fleet of EVs in a residential area. The simulation results show that the scheduled charging load can contribute to reducing peak loads by 23% at morning and 15% at evening hours. Also, the number of EVs, the initial SOC and battery capacity of EVs, with the smart scheduled charging and V2H and V2G technologies, was found to be proportionally related to the percentage of peak load reduction.

In the context of this simulation study, it was assumed that the EVs are connected only to the owners' houses; whereas in practice, the EVs could be connected the grid at any other location such as the destination of the trip, others' houses, etc. As future work, the proposed approach can be expanded and assessed while considering the mobility of the EV.

ORCID

Mouloud Denai  <https://orcid.org/0000-0001-8290-6743>

REFERENCES

1. Jordehi, A.R.: Optimisation of demand response in electric power systems, a review. *Renew Sustain. Energy Rev.* 103, 308–319 (2019)
2. Paterakis, N.G., Erdinç, O., Catalão, J.P.: An overview of demand response: key-elements and international experience. *Renew Sustain. Energy Rev.* 69, 871–891 (2017)
3. Datta, U., et al.: A price-regulated electric vehicle charge-discharge strategy for G2V, V2H, and V2G. *Int. J. Energy Res.* 43(2), 1032–1042 (2019)
4. Future Energy Scenarios (FES) Report. Published by UK's National Grid Operator, Warwick, UK (2018)
5. Vega-Fuentes, E., Denai, M.: Enhanced electric vehicle integration in the UK low-voltage networks with distributed phase shifting control. *IEEE Access.* 7, 46796–46807 (2019)
6. Datta, U., Kalam, A., Shi, J.: Electric vehicle (EV) in home energy management to reduce daily electricity costs of residential customer. *J. Sci. Ind. Res. (India).* 77(10), 559–565 (2018)
7. Chen, Y.W., Chang, J.M.: Fair demand response with electric vehicles for the cloud based energy management service. *IEEE Trans. Smart Grid.* 9(1), 458–468 (2016)
8. Li, Y., Li, K.: Incorporating demand response of electric vehicles in scheduling of isolated microgrids with renewables using a bi-level programming approach. *IEEE Access.* 7, 116256–116266 (2019)
9. Solanke, T.U., et al.: A review of strategic charging–discharging control of grid-connected electric vehicles. *J. Energy Storage.* 28, 101193 (2020)
10. Akhade, P., et al.: A Review on control strategies for integration of electric vehicles with power systems, In: *IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, pp. 1–5, Denver, (2018)
11. Wang, K., et al.: Distributed energy management for vehicle-to-grid networks. *IEEE network.* 31(2), 22–28 (2017)
12. Lazzaroni, P., et al.: Optimal battery management for vehicle-to-home and vehicle-to-grid operations in a residential case study. *Energy.* 175, 704–721 (2019)
13. Umetani, S., Fukushima, Y., Morita, H.: A linear programming based heuristic algorithm for charge and discharge scheduling of electric vehicles in a building energy management system. *Omega.* 67, 115–122 (2017)
14. Chiş, A., Lundén, J., Koivunen, V.: Reinforcement learning-based plug-in electric vehicle charging with forecasted price. *IEEE Trans. Veh. Technol.* 66(5), 3674–3684 (2016)
15. Najafi, S., et al.: Reinforcement learning method for plug-in electric vehicle bidding. *IET Smart Grid.* 2(4), 529–536 (2019)
16. Wan, Z., et al.: Model-free real-time EV charging scheduling based on deep reinforcement learning. *IEEE Trans. Smart Grid.* 10(5), 5246–5257 (2018)
17. Sayed, K., Gabbar, H.A.: Electric vehicle to power grid integration using three-phase three-level AC/DC converter and PI-fuzzy controller. *Energies.* 9(7), 532 (2016)
18. Habib, S., Kamran, M., Rashid, U.: Impact analysis of vehicle-to-grid technology and charging strategies of electric vehicles on distribution networks—a review. *J. Power Sources.* 277, 205–214 (2015)
19. Arnott, N.: UK Research and Innovation. [Online] (2018). <https://gtr.ukri.org/projects?ref=104248> Accessed Aug 2020
20. Manthey, N.: electrive.com. [Online] (2020). <https://www.electrive.com/2020/03/19/de-nissan-partners-complete-v2g-pilot-for-renewables/> Accessed Aug 2020
21. SEEV4-City, [Online] (2020). <https://www.seev4-city.eu/projects/> Accessed Aug 2020
22. Barnes, J.: Businesswire. <https://www.businesswire.com/news/home/20190305005307/en/Mitsubishi-Motors-World-Premiere-ENGELBERG-TOURER-New> (2019) Accessed Aug 2020
23. Schmidt, B.: The Driven [Online] (2019). <https://thedriven.io/2019/07/11/nissan-sees-leaf-as-home-energy-source-says-tesla-big-battery-waste-of-resources/> Accessed Aug 2020

How to cite this article: Alfaverh F, Denai M, Sun Y. Electrical vehicle grid integration for demand response in distribution networks using reinforcement learning. *IET Electr. Syst. Transp.* 2021; 1–14. <https://doi.org/10.1049/els2.12030>