Optimal Vehicle-to-Grid Control for Supplementary Frequency Regulation Using Deep Reinforcement Learning

Fayiz Alfaverh
School of Physics, Engineering and Computer Science
University of Hertfordshire, Hatfield, UK
fa17abh@herts.ac.uk

Mouloud Denaï
School of Physics, Engineering and Computer Science
University of Hertfordshire, Hatfield, UK
m.denai@herts.ac.uk

Yichuang Sun
School of Physics, Engineering and Computer Science
University of Hertfordshire, Hatfield, UK
y.sun@herts.ac.uk

Corresponding author: Mouloud Denai.

Abstract

The expanding Electric Vehicle (EV) market presents a new opportunity for electric vehicles to deliver a wide range of valuable grid services. Indeed, the emerging Vehicle-to-Grid (V2G) technology with bi-directional flow of power provides the grid with access to mobile energy storage for demand response, frequency regulation and balancing of the local distribution system by storing the surplus of electricity generated from intermittent renewable energy sources in EV batteries during off-peak periods and feeding it back to the grid when needed. This reduces electricity costs at peak hours and can be profitable for customers, network operators and energy retailers. In this paper, an optimal V2G control strategy using Deep Reinforcement Learning (DRL) is proposed to simultaneously maximise the benefits of EV owners and aggregators while fulfilling the driving needs of EV owners. In the proposed DRL-based V2G control strategy, a Deep Deterministic Policy Gradient (DDPG) agent is used for real-time decision-making. The DDPG dynamically adjusts the V2G power scheduling of the EV battery to satisfy the driving demand of EV users and simultaneously perform frequency regulation tasks dispatched by grid operator. The proposed V2G control scheme is tested on a power system consisting of two areas interconnected undergoing frequency deviations. The simulation results have shown that the proposed V2G control can lead to a 22.07% frequency deviation reduction and a 23.15% the Area Control Error (ACE), while satisfying the charging demands of EVs. The proposed V2G strategy has also been compared to other strategies to demonstrate its superior performance.

Keywords—vehicle-to-grid, frequency regulation, energy management, demand response, deep reinforcement learning.

1. Introduction

The raising penetration of intermittent Renewable Energy Sources (RESs) in the power grid introduces fluctuations in the generation side and this results in power mismatch between supply and demand leading to system frequency deviations. In addition, RESs such as wind power generators, are composed of power electronic converters, and when decoupled from the electricity grid they contribute to the reduction of the total power system inertia, thus affecting the overall dynamic stability of the system [1], [2].

Utility-scale Battery Energy Storage Systems (BESS) are a promising technology option to enhance the reliability and stability of the electricity grid and a key enabler for the transition towards a greener and reliable energy landscape. Owing to their rapid response time, BESSs are particularly well-suited for frequency regulation but can also provide other functions such as ramping, arbitrage and load following. Several recent studies have discussed the potential impact of BESS integration on the power grid's stability [3][4]. However, despite the

rapidly falling cost of BESS, utility-scale installations are currently not yet economically viable, but it is anticipated that there will a huge market opportunity over the next decades [5].

The electric vehicle market is growing very fast and is set to secure a record share of the global automotive market in the coming years. The principal drivers are the significant developments in batteries technologies, the economic and environmental benefits, and government's monetary incentives, such as tax exemptions or rebates to EV users [6]. According to the report of UK National Grid in Future Energy Scenarios [7], there will be 11 million of EVs by 2030 and 36 million by 2040.

Since the wide-scale adoption of EVs poses great challenges to the power system operation, namely increasing peak demand, stress on the transmission lines, and impacting the power system security. Therefore, the concept of Vehicle -to -Grid (V2G) was proposed to enable EVs to activity participate in the demand side management which aims to effectively contribute to frequency response services and improve the stability of power grid. V2G technology enables EVs to serve as a mobile battery storage device and with a bidirectional power flow, EVs can contribute to suppressing the frequency deviation by compensating the supply-demand mismatch caused by intermittent energy sources such as wind and solar energy [8]. In addition to the fast response of the EV battery, makes the EVs more attractive to provide various ancillary services such as frequency regulation [9].

Consequently, different V2G pilot projects have been carried out across the world over the past few years. These V2G technologies have also been adopted by the world's largest car manufacturers and are already in the marketplace. In the UK, Vehicle to Grid Britain (V2GB) project evaluates the long-term integration of V2G into the UK's utility network as well as the early opportunities for this technology in the UK power markets, offering services to the System Operator (SO), mainly Firm Frequency Response (FFR) [10]. In Italy, Fiat Chrysler Automobiles (FCA), in partnership with ENGIE, has started the first phase of its V2G pilot project which will use EVs' batteries to provide grid services [11].

The participation of EVs in the grid frequency regulation has been the subject of intensive research in recent years. In [12], the authors focused mainly on the EV contributions for primary frequency control to enable a secure and large-scale integration of intermittent renewable energy sources. In [13], the authors proposed a load frequency control (LFC) strategy based on V2G technology. The simulation results show how the V2G power control can be applied to compensate for the inadequate LFC capacity and thereby to improve the frequency stability of power grids. In [14], the stability of the LFC system in a microgrid with EVs and communication delay is investigated. In [15], the authors analysed the impact of the integration of EVs and RESs on the grid stability. A standard IEEE 13-bus test feeder is used in this study under a number of scenarios which are critical for the grid stability. In [16], the authors developed a droop-based control scheme to adjust the V2G power of the EV battery according to the frequency signal. A V2G control was proposed in [17] to enable EVs to actively participate in frequency regulation considering high frequency regulating signals. The authors in [18] developed a power model of EVs for effective frequency regulation considering wide-scale wind integration.

However, the storage capacity of a single EV's battery cannot provide frequency regulation services. Thus, a V2G aggregator agent is designated as a mediator between the utility network's operator and the fleet of EVs providing grid ancillary services. The V2G aggregator plays an important role in managing the charging and discharging of each EV participated in frequency regulation to ensure the satisfaction of EVs' driving demand and optimise the

tracking performance of frequency control signal [19]. Thus, the optimal dispatch strategy is crucial for the V2G aggregator to ensure that the EV driving needs are fulfilled while providing optimal frequency regulation to the power grid.

The dispatching strategy for V2G aggregator participating in frequency regulation has become a research hotspot. Some researchers focused on proposing dispatch strategies based on economy problems to optimise ~~the economy~~ the monetary benefits of EV owners or EV aggregators [20]-[22]. In [20], the authors assessed the economic profits of EVs participating in the frequency regulation markets from the perspective of the EV owner. The simulation results show that the EV owner can make an annual profit ranging between €100 and €1100 for participating in frequency regulation. In [21], a mixed integer linear programming (MILP) is used to optimise the daily benefits for EV users that can be gained from providing frequency regulation. In [22], the expected V2G incomes from participating in frequency regulation is estimated considering the EV battery price. The simulation results show that the estimated benefits exceed the EV battery prices in the current markets indicating that V2G regulation is an economically viable grid service. Nevertheless, these studies mainly focused on benefits and costs from the electricity market, where the regulation tasks and the charging demand of the battery were not considered.

Other studies have focused on the optimal dispatching strategies considering the capacity of EVs participating in frequency regulation. In [23], a queuing network model is used to predict the number of EVs to estimate the energy storage capacity required for frequency regulation based on a constant charging power for each EV. In [24], the authors proposed an optimal dispatch strategy for V2G aggregator to satisfy the driving demands of EVs and maximise the economic benefits of the aggregator while providing frequency regulation. However, the expected State-of-Charge (SOC) of the EV battery was considered as an inequality constraint under which the scheduled charging of EVs could not be performed by the proposed optimal strategy. In [25], a V2G control strategy is proposed to achieve the frequency regulation considering the expected SOC levels of EV batteries while providing real-time adjustments of their scheduled V2G power. However, these adjustments on the V2G power reduce the EV battery capacity to perform frequency regulation.

Several studies have focussed on the design of optimal V2G control strategies for EVs to maximise the frequency regulation capacity and maintaining SOC levels within the expected range [5][26]-[28]. An optimal dispatch strategy is proposed in [5] using the modern interior point optimisation method to enhance EVs contribution in SFR considering the driving needs of EV owners. The dispatch signal is fairly distributed from the control centre among EVs using the Area Control Error (ACE) and Area Regulation Requirement (ARR) criteria. In [26], a dynamic strategy for EV frequency regulation is proposed considering the driving patterns of EV owners and the EV charging/discharging power is regulated based on the frequency signal. The authors in [27] and [28] proposed a real-time V2G control using a state-space model which offers higher computational efficiency, higher accuracy, and a lower real-time communication requirement. Although all previously mentioned studies [5][26]-[28] have considered the charging demands of EVs participating in frequency response services, the underlying V2G control strategies are based on a forced charging process which allows EV batteries to charge/discharge quickly with maximum charging/discharging power rates before the plug-out time (disconnection). However, this leads to other issues for both the grid and the EVs' owners. For the power grid, forced charging may create a new EV charging peak demand, therefore imposing additional stress on the grid. This will also reduce the time during which EVs can offer frequency response services, since EVs will reach their maximum SOC and then become unable to

participate in frequency regulation. For EVs' users, the charging demands of EVs are not optimally satisfied, especially when ARR is dispatched for frequency regulation. In addition, EVs' owners cannot use their vehicles while they are participating in frequency regulation, even for their essential travel needs.

Reinforcement Learning (RL) is a machine learning, goal-oriented algorithm designed through a process of trial and error, and which strives to optimise for the actions which lead to the best rewards. RL is a very powerful approach which can be applied to almost any real-world system due to its ability to dynamically learn from an environment and discover possible actions. Deep Reinforcement Learning (DRL) combines the framework of RL with a multiple layer artificial neural network. DRL is most suitable for the optimal charging/discharging management of Electric Vehicle (EVs) connected to the electricity distribution grid. They have fast response and provide optimal and continuous control actions which is crucial for tracking the frequency of the power system in real-time. Multi-agent reinforcement learning has been proposed distributed electric vehicle charging coordination and fast V2G dispatch, considering the uncertainties and charging demands of EVs [29], 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 for V2G control considering the mobility of EVs, states of charge of EVs, and the estimated/actual [30]. However, these algorithms require historical data, such as driving patterns and battery State-of-Charge (SOC) as inputs to compute the charging/discharging schedules in real-time.

In this study, an optimal real-time V2G control strategy is designed for EVs to perform supplementary frequency regulation. The main feature that distinguishes the proposed approach from previous related works is that the scheduled charging power of an individual EV is optimally tracked and adjusted in real-time to fulfil the charging demand of EV's battery at the plug-out time without using the forced charging technique to maximise the frequency regulation capacity. The major contributions of this paper are summarized as follow:

- An optimal V2G control strategy based on both ACE and ARR signals is proposed for adjusting the scheduled charging power of EVs and maximising the frequency regulation periods without compromising EVs' users' preferences.
- Using Deep Reinforcement learning, a model-free V2G control is obtained which leads to a reduction in the computational time and hence an improved frequency response as compared to other studies.
- The proposed strategy has been tested under different EV charging/discharging scenarios.
- The proposed strategy is carefully evaluated in terms of improving the grid frequency response quality, reducing ACE signal of the system, and ensuring the expected driving demand of EV users.

The remaining of the paper is organised as follows: Section 2 presents an overview on frequency regulation in power systems and how EVs can participate in SFR. The proposed V2G control based on DDPG is illustrated in Section 3. Section 4 presents the simulation results and discussion. Finally, the conclusion of the paper is summarised in Section 5.

2. Electrical Vehicles Participation in Supplementary Frequency Regulation (SFR)

2.1 Overview of frequency response in power systems

The grid frequency indicates the balance between electricity supply and demand. If the total demand exceeds the total supply, then the frequency falls, while the frequency rises when the total supply is greater than the total demand. Thus, maintaining the frequency around the nominal value throughout the power system is critically important and requires a good control of the power output of the generator units in real time to ensure a reliable, secure and economic operation of the power grid. Therefore, the SFR seeks to keep the system frequency deviations within the normal limits.

Fig.1 presents the concept of SFR in the traditional power system. Where the Area Control Error (ACE) signal is generated as a weighted summation of the frequency deviation and the tie-line power changes and mostly belongs to a Gaussian type distribution with zero mean and fast switching between positive and negative values. Therefore, the SFR aims mainly to mitigate the ACE fluctuations as much as possible and maintain the frequency within an allowable limit by regulating the outputs of the generating units. While Area Regulation Requirement (ARR) refers to the supply-demand mismatch that needs to be restored by the generator units and generally remains positive or negative for a long period of time.
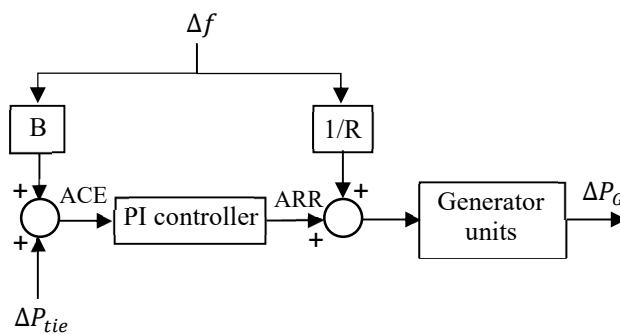

Fig. 1 SFR in the traditional power system.

2.2 Participation of EVs in SFR

EVs with V2G technology have the great ability to respond much faster to frequency response services than conventional generation units. Furthermore, EVs are usually utilised for about only 4% of the day and are idle (parked at home or at the workplace) for the rest of time.

When EVs participate in SFR, ACE and ARR signals become readily available to be dispatched for V2G control as illustrated in Fig. 2(a) and Fig. 2(b) respectively. However, there are some differences between ACE and ARR which must be considered when they are dispatched to EVs. Since ACE has almost a zero-mean distribution, thus the average value of ACE is around zero and hence the charging demand of EVs may not be satisfied. Therefore, using ACE for frequency regulation will not affect EVs' batteries SOC levels. Conversely, when ARR is dispatched for frequency regulation, the EVs' batteries SOC levels will deviate from the initial values since ARR signal may have positive/negative mean. Consequently, EVs may lose their capacity for regulation which impacts the frequency stabilisation. Therefore, each of the ACE and ARR signals has its advantages and disadvantages when used for frequency regulation.
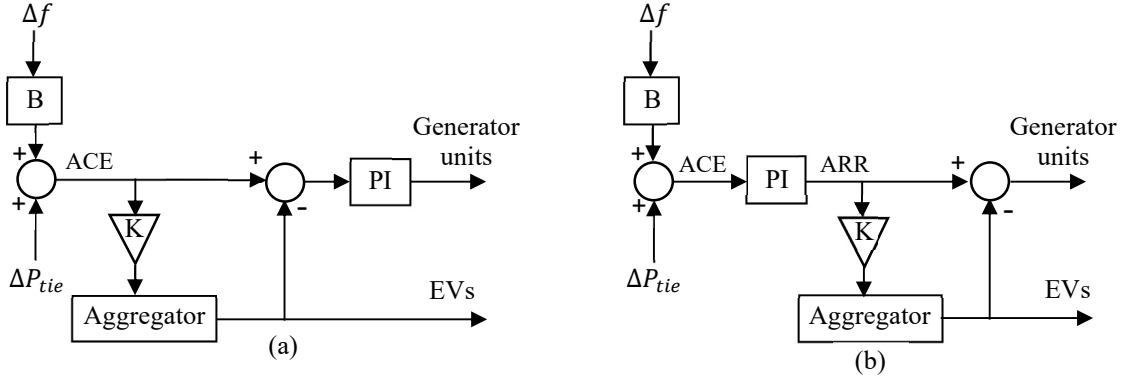
Fig. 2 ACE and ARR signals for regulation (a) ACE, (b) ARR.

2.3 Dispatch Strategies of EVs Participating in SFR

Recently with the emerging V2G technology, EV dispatching strategies for SFR have been the subject of extensive research in the literature. Fig. 3 illustrates the hierarchical dispatch strategy which has three levels: Control Centre, EV Aggregator and EVs fleet. Since the V2G power of a single EV battery is not large enough, mostly around 3 kW to 10 kW, EVs must be aggregated to participate in SFR.
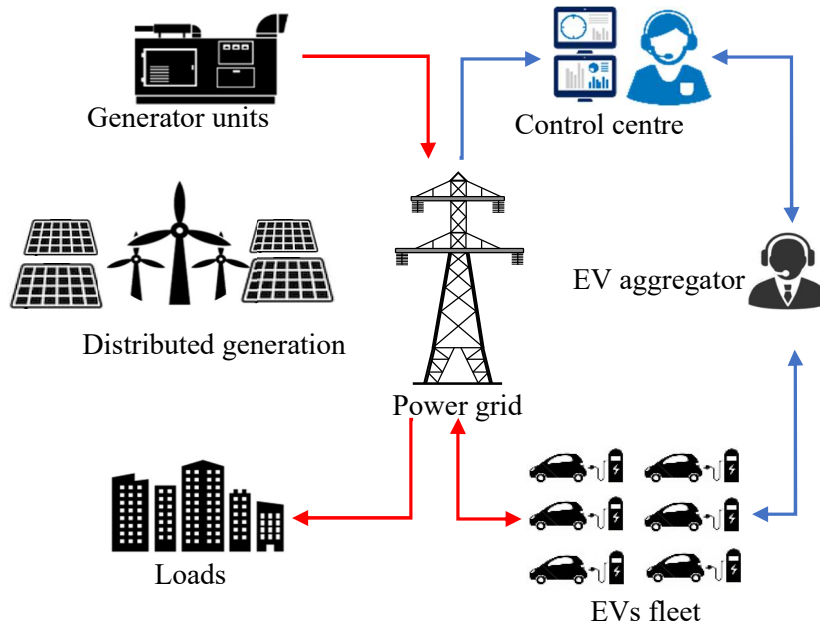

Fig. 3 Hierarchical control of EVs in the power system.

A. Dispatch strategy in the control centre

To minimise the energy mismatch between demand and supply as much as possible, the regulation signals (ACE or ARR) are dispatched to the generating units and EVs performing frequency regulation. While the regulation dispatch to EVs depends on many factors, such as the RES generation, the capacity of the generator units and the frequency regulation capacity (FRC) of EVs. Therefore, the dispatch regulation task to an EV aggregator from the control centre is formulated as:

$$P_t^{AG} = K \times \begin{cases} \max(S_t, -C_t^{up}, \Delta f \le 0) \\ \min(S_t, C_t^{down}, \Delta f > 0) \end{cases} \tag{1}$$

Where S_t is either the ACE or ARR signal at time t, P_t^{AG} indicates the dispatch regulation task undertaken via an individual EV aggregator, K is a ratio that determines the proportion dispatch based on which the EVs aggregator undertakes frequency regulation and takes a value between 0 and 1. C_t^{up} and C_t^{down} represent the total capacity of regulation-up and -down uploaded by the EV aggregator to the control centre respectively, which can be calculated by adding up the FRC of all EV charging stations ($C_{j,t}^{up}, C_{j,t}^{down}$) as follows:

$$\begin{cases} C_t^{up} = \sum_{j=1}^{J} C_{j,t}^{up} \\ C_t^{down} = \sum_{j=1}^{J} C_{j,t}^{down} \end{cases} \tag{2}$$

B. Dispatch strategy in the EV aggregator

The role of the aggregator is to facilitate the integration of EVs in the electricity market. The aggregator agent acts as a commercial middleman or intermediary between EVs and the power grid. The aggregator receives the frequency control signal from the control centre and sends the power capacity collected from EVs back to the control centre. The aggregator then dispatches the charging or discharging commands to each EV considering the frequency control signals. Therefore, the grid operator (i.e. control centre) communicates with the aggregator agents only, and it does not need to manage each individual EV, as shown in Fig. 2.

Once the regulation power signal reaches the EV aggregator from the control centre, the regulation power must be distributed proportionally among the charging stations based on their corresponding FRCs uploaded by the aggregator, the regulation power of each charging station is defined as follow:

$$P_{j,t}^{st} = \begin{cases} P_t^{AG} \times \dfrac{C_{j,t}^{up}}{C_t^{up}}, & P_t^{AG} \le 0 \\ P_t^{AG} \times \dfrac{C_{j,t}^{down}}{C_t^{down}}, & P_t^{AG} > 0 \end{cases} \tag{3}$$

Where $P_{j,t}^{st}$ denotes the power regulation that needs to be provided by the jth EV charging station at time t.

C. Dispatch strategy in EV charging stations

In the charging station, the regulation power will be distributed to each EV based on the amount of FRC achieved in the previous time step:

$$\Delta P_{i,t}^{reg} = \begin{cases} P_{j,t}^{st} \times \dfrac{C_{i,t}^{up}}{C_{j,t}^{up}}, & P_{j,t}^{st} \le 0 \\ P_{j,t}^{st} \times \dfrac{C_{i,t}^{down}}{C_{j,t}^{down}}, & P_{j,t}^{st} > 0 \end{cases} \tag{4}$$

Where $\Delta P_{i,t}^{reg}$ denotes the power deviation for the regulation of the ith EV at time t and $C_{j,t}^{up}/C_{j,t}^{down}$ is the total capacity regulation-up/regulation-down of the charging station, and can be calculated as follows:

$$\begin{cases} C_{j,t}^{up} = \sum\nolimits_{i=1}^{I} C_{i,t}^{up} \\ C_{j,t}^{down} = \sum\nolimits_{i=1}^{I} C_{i,t}^{down} \end{cases} \qquad (5)$$

Where

$$\begin{cases} C_{i,t}^{up} = P_{max} + P_{i,t} \\ C_{i,t}^{down} = P_{max} - P_{i,t} \end{cases} \qquad (6)$$

When the V2G power is positive (negative) as shown in Fig. 4, the FRC of regulation-up is higher (lower) than the FRC of regulation-down. Therefore, the regulation task, whether regulation-up or regulation-down, depends mainly on the corresponding FRC. This can lead to a deviation in the SOC of the EV battery from the expected level. To this end, the V2G power must be continuously regulated to ensure that the expected charging demands of EVs owners are satisfied while simultaneously providing frequency regulation services.
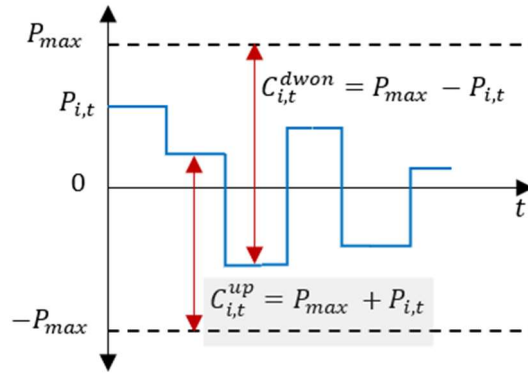

Fig. 4 Available frequency regulation capacity of an individual EV.

3. Deep Reinforcement Learning for V2G Control

Reinforcement Learning (RL) is a machine-learning technique for optimal decision-making in a stochastic environment. In RL, at each step t, an agent interacts with the environment by executing an action $a_k \in \mathcal{A}$ based on a given policy $\pi: \mathcal{S} \to \mathcal{A}$ at a present state $s_k \in \mathcal{S}$. The new state s_{k+1} is then computed and the agent receives a numerical reward $r_k \in \mathcal{R}$ on the action taken that generated by a reward function of both action and state $r_k(s_k, a_k)$. The probability function $p(s_{k+1}, r_k | s_k, a_k)$ is used to model the transition between the states. The agent strives to optimise the policy and hence maximise the expected long-term reward $G_k = \sum_{i=0}^{\infty} \gamma^i r_{k+1}$ at each time step t, where $\gamma \in [0,1]$, represents the discount factor. The Q-function refers to the quality of the action taken in each state and is used in many RL algorithms to find the optimal policy and is defined by:

$$Q^{\pi}(s, a) = \mathbb{E}_{\pi}[G_k | s_k = s, a_k = a] \qquad (7)$$

Where $Q^{\pi}(s, a)$ denotes the expected discounted return for taking an action a in a certain state s and under a given policy π.

Deep Reinforcement Learning (DRL) aims mainly to optimise the policy using deep neural networks by approximating the action value functions and optimal policy as shown in Fig. 5 [31]. One of the most common deep RL algorithms is Deep Deterministic Policy Gradient (DDPG) due to its ability of solving continuous state and action optimisation problems.
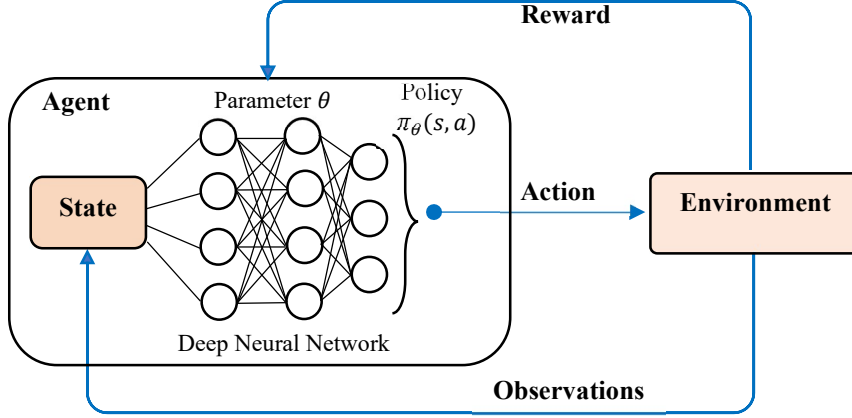
Fig. 5 Deep reinforcement learning process [31].

3.1 Deep Deterministic Policy Gradient (DDPG)

Using the DDPG agent, two neural networks are employed to approximate the Q-function given by equation (7) and the policy function π. These neural networks are known as critic network $Q(s_k, a_k|\theta^Q)$ and actor network $\mu(s_k|\theta^\mu)$, with θ^Q and θ^μ being parameters of these networks, respectively. For a certain state, the actor network will take an action as an output. Then the state-action pair will be transmitted to the critic network as an input returning the Q-value based on the state-action pair. In addition, two other networks, namely, the target actor network $Q'(s, a|\theta^{Q'})$ and the target critic network $\mu'(s|\theta^{\mu'})$ are used to continuously update the parameters of the actor and critic networks, which significantly improve the stability of the optimization. Under these conditions, the target networks' parameters ($\theta^{Q'}$ and $\theta^{\mu'}$) are used to update the traditional networks' parameters (θ^Q and θ^μ) at every time step using the smoothing factor τ as follows:

$$\begin{cases} \theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \\ \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \end{cases} \tag{8}$$

To let the agent explores the environment and interact with it during the training process, the number of episodes must be defined where each episode consists of a series of steps. After each iteration, a sampled noise $N(\sigma)$ generated from the Ornstein-Uhlenbeck (OU) process is added to the output of the actor network: $a_k = \mu(s_k|\theta^\mu) + \mathcal{N}(\sigma)$, while the hyperparameter σ is utilised to evaluate the exploring process of the environment. Then the experience replay memory $\mathcal{D}$ (experience buffer) is used to store the tuple (s_k, a_k, r_k, s_{k+1}), and take the minibatches (independent and identically distributed sets of samples) for training. The critic network will be updated by minimising the loss function across all selected experiences (N):

$$L(\theta^Q) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - Q(s_i, a_i|\theta^Q)\right)^2 \tag{9}$$

Where the sample y_i is determined from the sum of the present reward and the expected Q-value at the next state s_{i+1} considering the outputs of target actor and critic networks:

$$y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'\left(s_{i+1}|\theta^{\mu'}\right)|\theta^{Q'}\right) \tag{10}$$

To improve the policy, the score function is minimized by using the gradient ascent to the actor network. The gradient is approximated by averaging the gradients of the policy score function across the minibatch:

$$\nabla_{\theta\mu}J \approx \frac{1}{N}\sum_{i=1}^{N}\left(\nabla_a Q(s_i, a|\theta^Q)|_{a=\mu(s_i)}\nabla_{\theta\mu}\mu(s_i|\theta^\mu)\right) \tag{11}$$

3.2 V2G Control Based on DDPG

The uncertain dispatches (ACE or ARR) from the control centre can be different since EVs have different FRCs. Consequently, this results in a change of the battery SOC from the expected level. Therefore, the Scheduled Charging/discharging Power (SCP) should be continuously regulated in real-time to satisfy the EV's charging demand.

However, when regulation-up is dispatched, the SCP needs to be increased for the next step, leading to an increase of the real-time V2G power. Consequently, the FRC of the EV will increase for regulation-up and will decrease for regulation-down. Conversely, when the SCP decreases due to performing regulation-down, the FRC of the EV will decrease/increase for regulation-up/regulation-down. When the SCP reaches the maximum V2G power, EVs cannot perform SFR since, in this case, they have no capacity for regulation-down and performing regulation-up leads to a further loss in the battery SOC. Thus, in such a scenario, the only option is to charge the EV battery. Similarly, if the SCP reaches the negative maximum of V2G power, the EV battery is unable to perform SFR either.

Normally, when an EV is not participating in SFC, the expected battery charging level can be achieved by charging the battery under the expected charging power (ECP). Thus, the expected SOC of the EV battery is defined as:

$$SOC_{i,k}^{exp}\big|_{P_i^{ecp}} = SOC_i^{init} + \frac{\Delta k . P_i^{ecp}}{E_i^{rat}} \tag{12}$$

Where

$$P_i^{ecp} = \frac{(SOC_i^{exp} - SOC_i^{init})E_i^{rat}}{t^{out} - t^{in}} \tag{13}$$

$SOC_{i,k}^{exp}\big|_{P_i^{ecp}}$ denotes the real-time expected battery SOC undertaking only ECP, E_i^{rat} is the capacity of the ith EV battery. SOC_i^{exp} and SOC_i^{init} are the predefined expected SOC and initial SOC of ith EV, respectively, Δk represents the time step for adjusting SCP, t^{out} and t^{in} are the plug-out and plug-in times, respectively.

However, when the EV is performing SFC, its battery's SOC of its battery will change based on the regulation undertaken and the scheduled charging power and can be calculated in real time as:

$$SOC_{i,k}\big|_{P_i^{schd}, P_i^{reg}} = SOC_i^{init} + \frac{\Delta t . \Delta P_{i,t}^{reg} + \Delta k . P_{i,k}^{schd}}{E_i^{rat}} \tag{14}$$

Where $SOC_{i,k}\big|_{P_i^{schd}, P_i^{reg}}$ is the actual battery SOC undertaking regulation and SCP, $P_{i,k}^{schd}$ represents the SCP at time k and Δt is the time step for regulation.

In order to achieve the charging demand of EV participating frequency regulation at the plug-out time without the need for forced charging process, the expected battery SOC ($SOC_{i,k}^{exp}\big|_{P_i^{ecp}}$) and the actual battery level ($SOC_{i,k}\big|_{P_i^{schd}, P_i^{reg}}$) should be tracked in real-time. The SCP is then continuously adjusted to minimise the difference between the actual and expected SOC of the EV battery.

Therefore, the objective function of the optimal V2G closed-loop control is a function of the change in SCP ($\Delta P_{i,k}^{schd}$) and can be expressed as:

$$\min f\left(\Delta P_{i,k}^{schd}\right) = \left|\frac{\Delta t.\Delta P_{i,t}^{reg} + \Delta k.(P_{i,k-1}^{schd} + \Delta P_{i,k}^{schd}) - \Delta k.P_i^{ecp}}{E_i^{rat}}\right| \qquad (15)$$

Where $\Delta P_{i,k}^{schd}$ is the change in SCP of the ith EV at time k and $P_{i,k-1}^{schd}$ is the SCP at time $k-1$.

Therefore, the objective of the DRL-based V2G control agent is to provide optimal actions to minimise the objective function given by equation (15). The DDPG algorithm employing the actor-critic approach is used to solve the Q-function of equation (7), The actor network generates a V2G control action, where the quality of the action taken will be evaluated by the critic network. The actor network takes the vector state s_k of $SOC_{i,k}\big|_{P_i^{schd},P_i^{reg}}$, difference between the actual and expected SOCs of the EV battery as an error e_k, and its integral $\int e_k.dk$ as input, and directly generates a continuous action a_k as the deviation of SCP ($\Delta P_{i,k}^{schd}$). The critic network, on the other hand, receives the state s_k and the action $\mu(s_k|\theta^\mu)$ as input and produces a scalar Q-value ($Q(s_k,a_k|\theta^Q)$) as shown in Fig. 6.
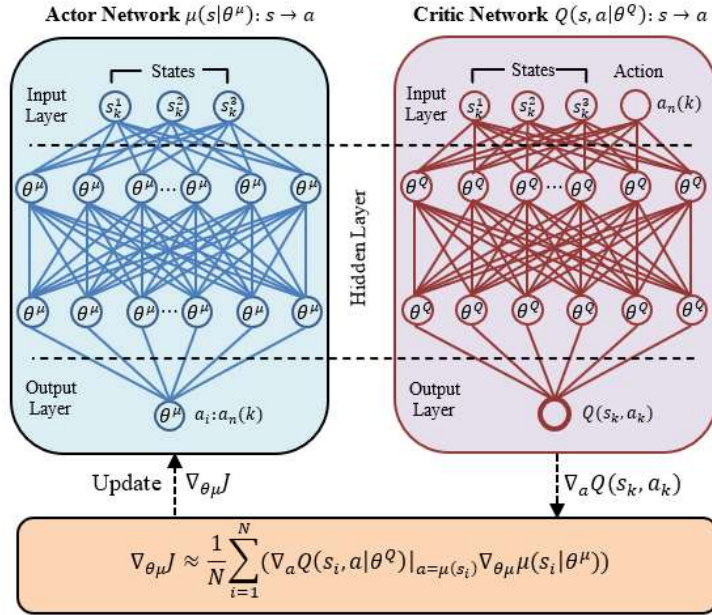

Fig. 6 Architecture of the actor and critic networks.

To evaluate the performance of the V2G control with the accumulated reward, the reward r_k in the DDPG algorithm is considered as:

$$r_k = \begin{cases} \text{Positive reward } (+R_p) \ \forall \ |e_k| \in [0, 0.05] \\ \text{Negative reward } (-R_n) \ \forall \ |e_k| \notin [0, 0.05] \\ \text{Large penalty } (-R_e) \quad \forall \ SOC_{i,k}\big|_{P_i^{schd},P_i^{reg}} \in [20, 80] \end{cases} \qquad (16)$$

When training the DDPG agent, M episodes will be repeated, while each episode consists K steps corresponding to the instants at which the agent-environment interactions take place. The training episodes or scenarios are created by selecting random variables to initialise the environment. The training procedure of the DDPG agent is described by the pseudo-code presented in Table 1.

Table 1 Pseudo-code of the DDPG algorithm.

1: Initialise critic $Q(s_k, a_k|\theta^Q)$ and actor $\mu(s_k|\theta^\mu)$ networks.
2: Initialise target networks $Q'(s, a|\theta^{Q'})$ and $\mu'(s|\theta^{\mu'})$ with traditional networks' parameters θ^Q and θ^μ
3: Set up empty reply buffer $\mathcal{D}$.
4: **for** episode $= 1, 2, \cdots, M$, **do**
5:     Generate random values for the environment's variables.
6:     Initialise the environment by simulating generated variables.
7:     observe the initial state variables.
8:     **for** $k = 1, 2, \cdots, K$, **do**
9:       Select the action using $a_k = \mu(s_k|\theta^\mu) + \mathcal{N}(\sigma)$
10:      Take action and receive the immediate reward r_k and the next state s_{k+1}.
11:      Store tuples (s_k, a_k, r_k, s_{k+1}) in $\mathcal{D}$.
12:      Sample a random minibatch of tuples from $\mathcal{D}$.
13:      set $y_i = r_i + \gamma Q'\big(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}\big)$
14:      Update critic network with the loss function:
$$L(\theta^Q) = \frac{1}{N} \sum_{i=1}^{N} \big(y_i - Q(s_i, a_i|\theta^Q)\big)^2$$
15:      Update actor networks using sampled policy gradient:
$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_{i=1}^{N} \Big(\nabla_a Q(s_i, a|\theta^Q)|_{a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s_i|\theta^\mu)\Big)$$
16:      Update target networks using:
$$\begin{cases} \theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \\ \theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \end{cases}$$
17:     **end for**
18: **end for**

4. Simulation Results and Discussion

The two-area interconnected power system and the thermal power plant model used in this simulation study, adapted from [26][32][33], are shown in Figs.7 and 8 respectively. The parameter values are listed in Table 2 [26]. The EVs are assumed to be connected to Area A. The random load and wind turbine power deviations follow a normal distribution with the zero mean as presented in Figs. 9 and 10, respectively.

Table 2 Parameters of two areas power system [26].

| Parameters	Area A	Area B
Maximum load capacity (MW)	20,000	10,000
Proportional and integral gains	10,0.01	1,0.01
Frequency regulation sample time (s)	4	--
Scheduled charging power sample time(s)	60	--
Frequency bias factor (pu/Hz)	0.15	0.075
Inertia constant (pu. s)	0.32	0.16
Load damping coefficient (pu/Hz)	0.04	0.02
Dead band of primary frequency detection (s)	0.033	0.033
Communication delay (s)	1	1
Dead band of ACE (MW)	10	10
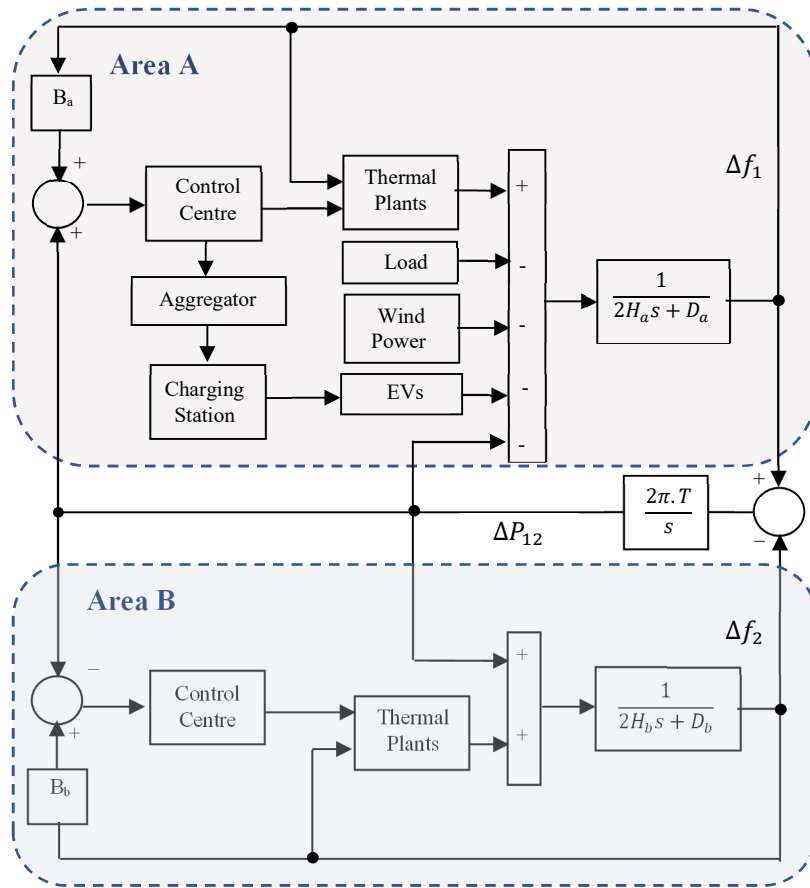
Fig. 7 Two-area power system model
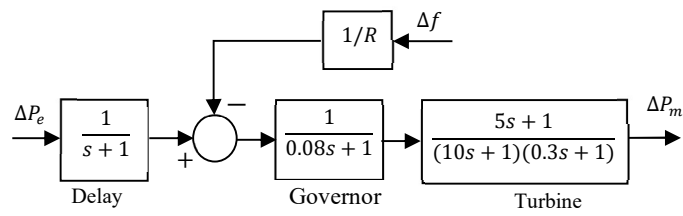
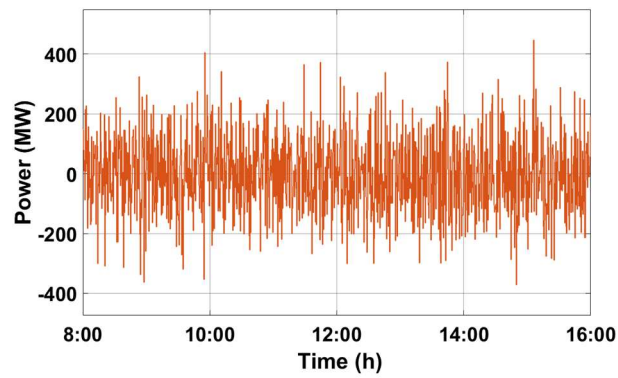Fig. 8 Thermal power generator for frequency control [26].

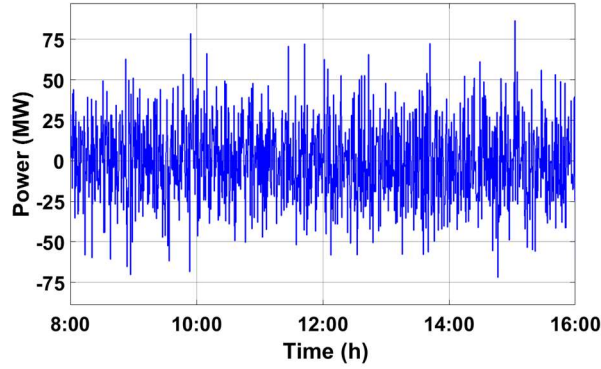Fig. 9 Load deviation in Area A.

Fig. 10 Random wind turbine power fluctuations.

In this simulation study, the EV aggregator is assumed to manage 100 EV charging stations and each station accommodates 500 EVs, as illustrated in Table 3. The EVs initial SOC, the expected SOC, arriving time, and departure time are generated randomly using Monte Carlo simulation. Since the expected SOC levels depend on the driving behaviour of the EV owners, thus in this paper, EVs are categorised into three types; Type I refers to those EVs that need to be charged, Type II are EVs that need to be discharged (selling energy), and Type III refers to EVs whose owners are not interested in charging from the utility grid or selling energy to the grid. The other parameters associated to EVs' batteries, such as rated capacity, higher/lower limit of SOC, and maximum charging/discharging power are also shown in Table 3.

Table 3 EVs Parameters used in the simulations.

Parameters	
Total number of charging stations	100
Total number of EVs in charging station	500
Number of EVs of Type I	350
Number of EVs of Type II	100
Number of EVs of Type III	50
Arriving time (h)	Time ~ N (9,0.1)
Departure time (h)	Time ~ N (16,0.1)
Initial SOC for Type I	SOC ~ N (0.4,0.01), SOC $\in$ [0.3,0.5]
Initial SOC for Type II	SOC ~ N (0.7,0.01), SOC $\in$ [0.6,0.8]
Initial SOC for Type III	SOC ~ N (0.6,0.01), SOC $\in$ [0.5,0.8]
Expected SOC for Type I	SOC ~ N (0.7,0.01), SOC $\in$ [0.6,0.8]
Expected SOC for Type II	SOC ~ N (0.4,0.01), SOC $\in$ [0.3,0.5]
Rated capacity of EV battery (kWh)	[24, 28, 32, 36, 40]
Rated charging and discharging power (kW)	7
Charging/discharging efficiency	0.9
SOC of EV battery limits (p.u)	0.8/0.2

To evaluate the effectiveness of the proposed dispatch strategy denoted as CS3, the results have been compared to two dispatching strategies from the literature. The first dispatch strategy is proposed in [33] and is termed CS1, which adjusts the SCP of the EV hourly to achieve the expected SOC level while participating in frequency
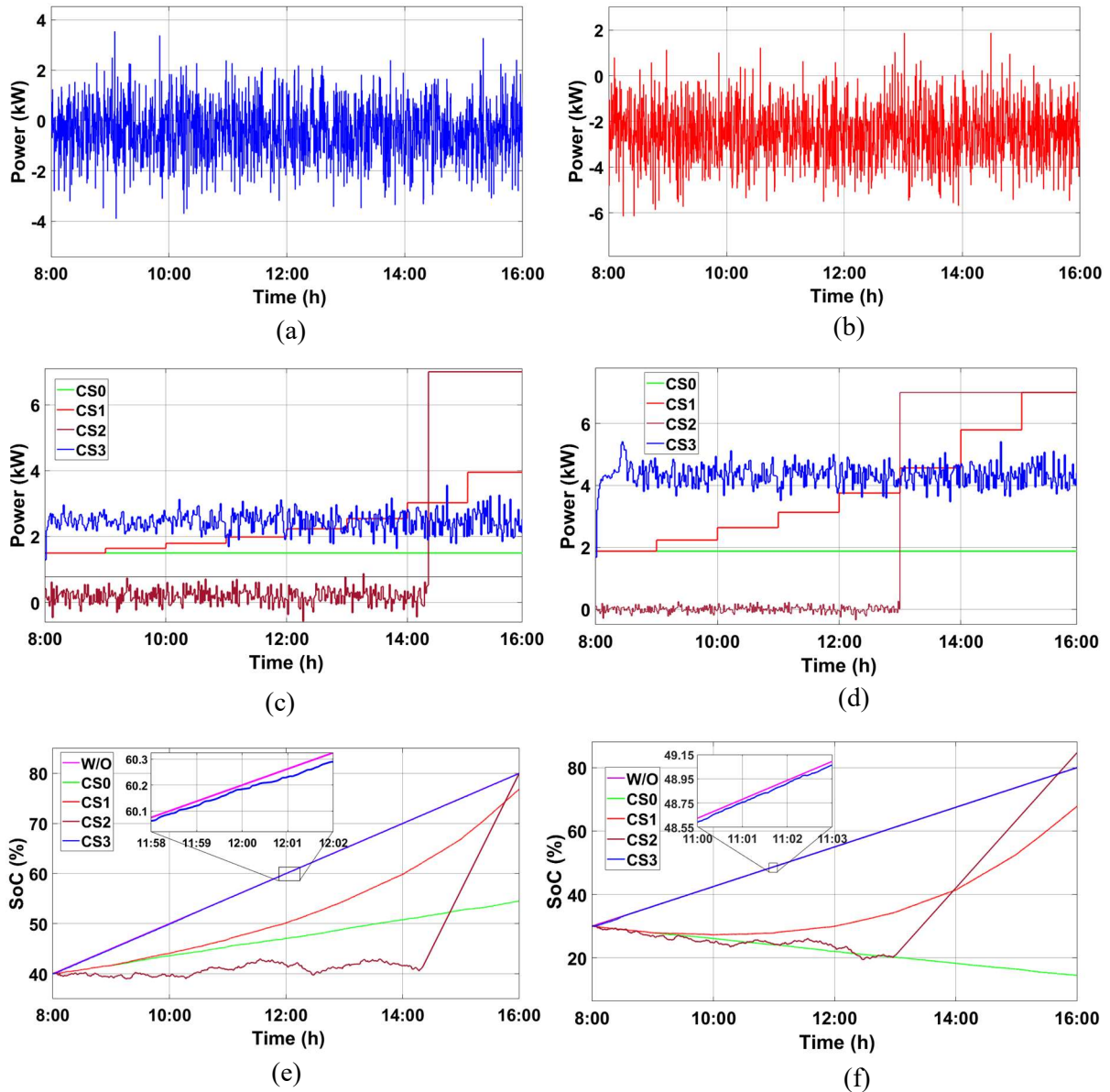
Fig. 11 Implementation of ACE and ARR signals on EVs of Type I; (a) ACE dispatch (b) ARR dispatch (c) SCP of Type I under ACE (d) SCP of type I under ARR (e) SOC battery of Type I under ACE (f) SOC battery of Type I under ARR.

regulation. The second dispatch strategy is proposed in [26] (referred to as CS2), which aims to satisfy the driving demand of EVs' owners while performing frequency regulation using the forced-charging method.

A. Impacts of the proposed V2G strategy on the EV battery

To evaluate the robustness of the proposed V2G control strategy, the dispatch signals ACE and AAR received from the control centre and implemented on the chosen EVs are simulated with large fluctuations and large mean values respectively and the results are compared to other strategies.

For EVs of Type I, the owners chose to charge their EVs' batteries to the upper SOC level. Therefore, the proposed V2G strategy has been tested on this type of EVs by dispatching regulation using ACE with zero-mean and higher amplitude, and large negative-mean of ARR as shown in Fig.11(a) and Fig.11(b), respectively.

Fig.11 (c) and Fig.11 (d) show the real-time SCP adjustments for the EV to satisfy the charging demand. Under CS0 (basic strategy), the SCP of the EV remains constant all the time while the EV is participating in frequency
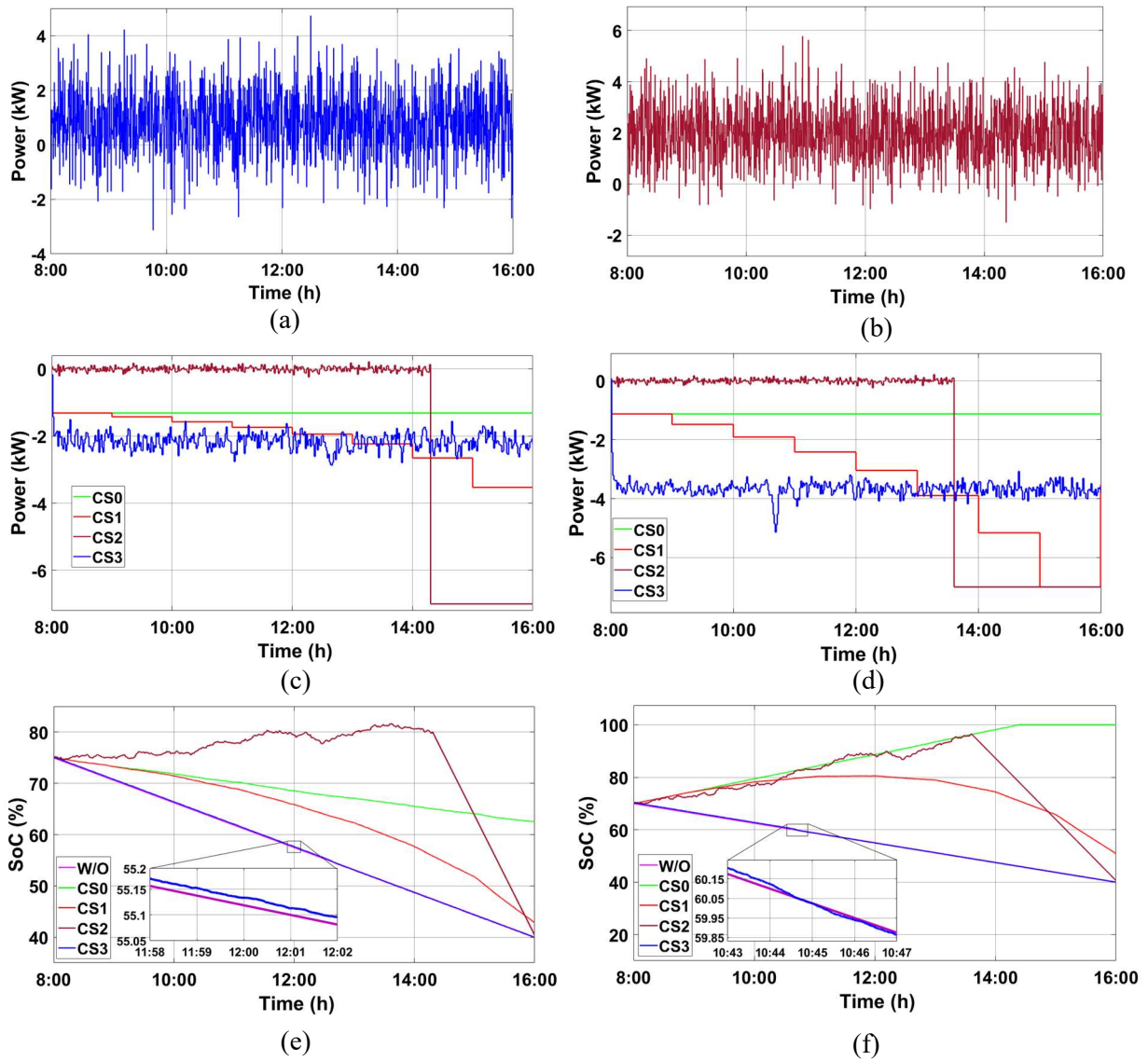
Fig. 12 Implementation of ACE and ARR signals on EVs of Type II; (a) ACE dispatch (b) ARR dispatch (c) SCP of Type II under ACE (d) SCP of type II under ARR (e) SOC battery of Type II under ACE (f) SOC battery of Type II under ARR.

regulation. This leads to a poor performance while trying to fulfil the expected charging demand of the EV before the plug-out time as shown in Fig.11 (e) and Fig.11 (f). While using CS1, the EV is losing its capacity for regulation-down, and this is because of the increase in SCP at each hour to compensate for the accumulated decrement of the EV's battery storage due to a large negative mean of ARR, even though it is not able to achieve the desired SOC. The forced charging method is used in CS2, where the EV can participate in the frequency regulation as needed and then charging at the maximum rate (7 kWh) to reach the expected SOC. Although the use of this technique can satisfy the driving demand of the EV, but it has some disadvantages. During the period of charging at the maximum rate, the EV will lose its capacity for the regulation-down leading to a poor performance in the quality of EVs' participation in frequency regulation. On the other hand, the forced charging may cause a new EV charging peak demand, therefore imposing additional stress on the grid. In contrast to CS2, CS3 can keep the EV's capacity for regulation up and down throughout the plug-in period by continuously adjusting the SCP while considering the expected charging demand of the EV at plug-out time.
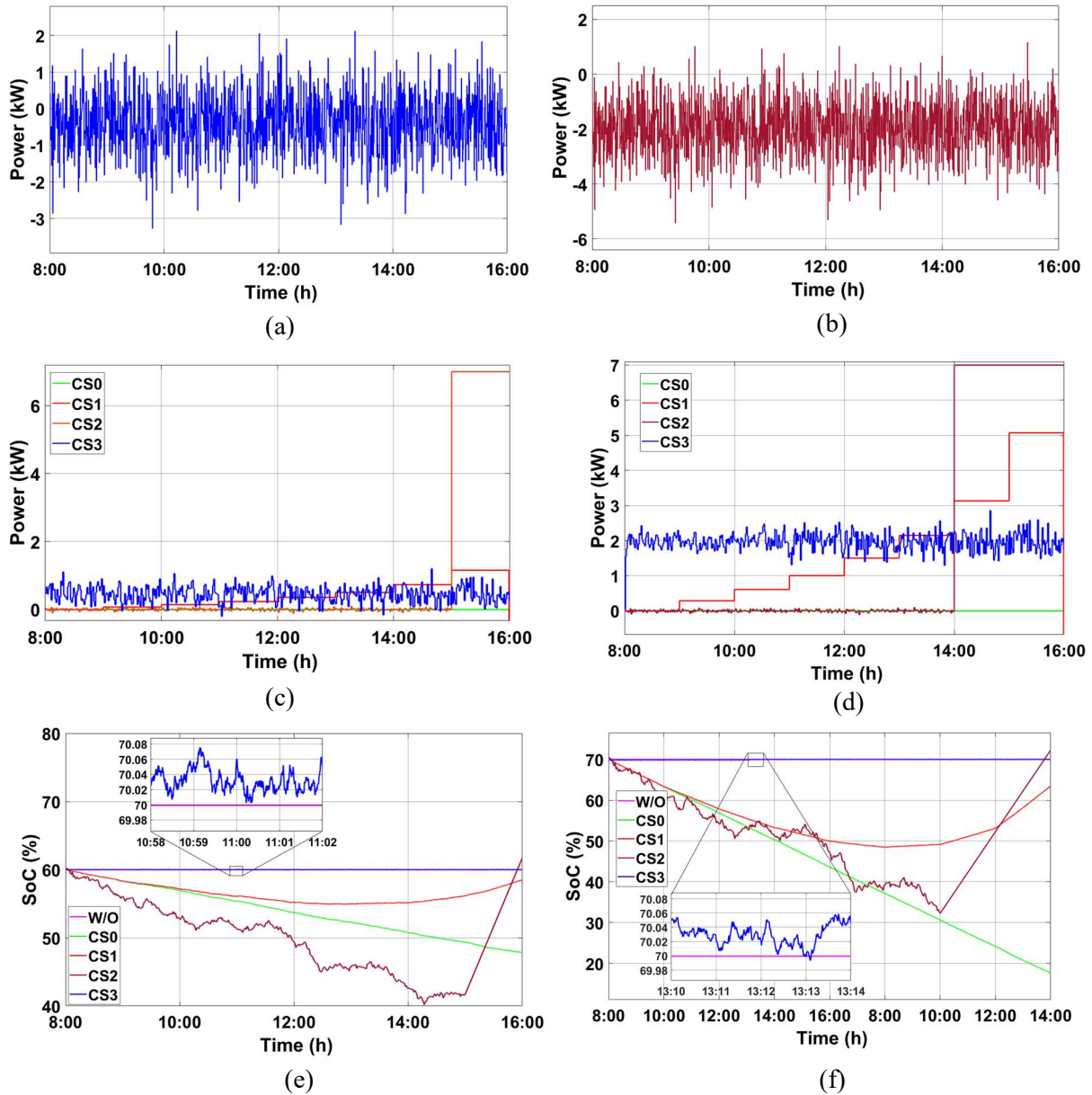
Fig. 13 Implementation of ACE and ARR signals on EVs of Type III; (a) ACE dispatch (b) ARR dispatch (c) SCP of Type III under ACE (d) SCP of type III under ARR (e) SOC battery of Type III under ACE (f) SOC battery of Type III under ARR.

For Type II, EV owners may prefer to trade the extra energy stored in the EV battery with the utility grid, especially when the SOC is high enough for travel needs. Therefore, this type of EVs may choose to discharge the battery's stored energy to the predefined SOC level. The dispatched regulation tasks using ACE and ARR are presented in Figs. 12 (a) and Fig. 12(b), respectively.

It can be observed that CS3 has advantages over CS0 and CS1 as it can participate in the regulation until the departure time due to its ability to keep the capacity for regulation-up as shown in Fig. 12 (c) and Fig. 12(d). In addition, as shown in Fig. 12 (e) and Fig. 12 (f), CS3 can guarantee the satisfaction of the charging demand, while CS0 and CS1 cannot fulfil the expected SOC levels when a large negative mean of ARR is dispatched to the EV. Although CS2 can optimally achieve the expected SOC of the EV at the plug-out time as shown in Fig.12 (e) and (f), however by using this strategy, the EV cannot participate in regulation-up for some time due to the maximum discharging rate as shown in Fig.12 (c) and Fig.12 (d).

When EV owners have enough SOC in their EVs' battery for a next trip, they may prefer to keep their battery at the same SOC level. However, EVs can participate in SFC during plug-in time provided that the same initial SOC is returned at plug-out time. The ACE and ARR dispatched to this type of EV are shown in Fig. 13 (a) and Fig. 13 (b). It can be noted that the optimal strategy (CS3) leads to a better performance as compared to all CS0, CS1 and CS2 strategies considering the large fluctuations in ACE and large mean of ARR as illustrated in Fig. 13 (c) and Fig. 13 (d). Besides achieving the regulation task, CS2 has good ability to minimise the difference between the real-time and the expected SOC as shown in Fig. 13 (e) and Fig. 13 (f).

B.   *Impacts of EVs on grid frequency regulation*

The frequency response capacity from EVs is influenced by the travelling time. The driving behaviour of EV owners working from 9:00 am to 16:00 pm is likely to follow the same pattern every day. Therefore, a normal distribution is considered for the arriving and departure times of EVs as shown in Table 3. Due to this distribution of times, the total FRC of EVs, including regulation up/down capacities is affected by the number of the EV arrivals to the charging stations. As expected, the FRC will start increasing from zero and reach the maximum capacity at the arriving time according to the number of EV arrivals and decreasing from the maximum capacity to zero at the departure time according to the number of EV departures from the charging stations as shown in Fig.14. The FCR of regulation-up and regulation-down will be uploaded in real-time to the control centre to dispatch the regulation tasks to EVs accordingly.
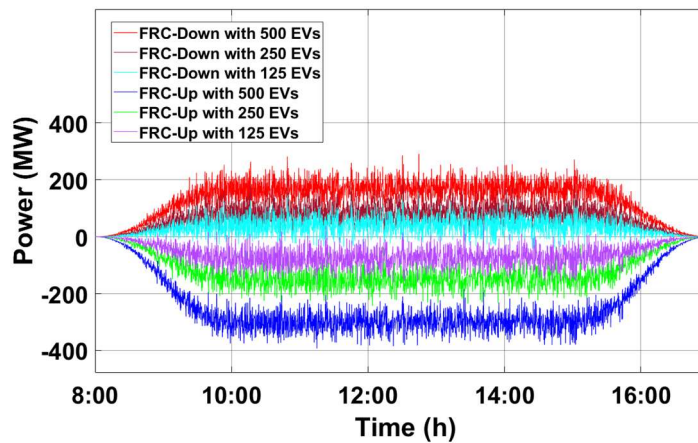

Figure 14 FRC of regulation-up and regulation-down under
different number of EVs in the charging station.

The total number of EVs parked at the charging stations and able to participate in frequency regulation significantly influences the frequency response capacity. To assess the impact of EVs' number to be used for frequency regulation, the model has tested under different sizes of EV fleets. Fig. 14 shows the frequency capacity for regulation-up and regulation-down with different number of EVs participating in the frequency regulation. It is noted that when the number of EVs has been reduced to half (i.e. 250 EVs), the FRC has been reduced accordingly, leading to a reduction in the capability of EVs for participating in frequency regulation. Therefore, the larger the number of EVs participating in frequency regulation leads to more FRC for regulation-up and regulation-down, and therefore a better frequency regulation service.

The dispatch strategy used for EVs to achieve the expected SOC can also influence the FRC. While considering CS1, CS2 and CS3, the regulation capacity of EVs under these strategies is analysed and presented in Fig.15. It can be observed that EVs are gradually losing their FRC of regulation up and regulation down under CS1, and this reduction in FRC occurred because of the continuous adjusting SCP every hour to reach the EV's charging demand at the departure time. It is also noted that the FRC under CS2 starts decreasing at 13:00 pm, this is due to the forced charging used to satisfy the expected SOC of the EVs after participating in frequency regulation, leading to the inability of the EVs to perform regulation-up and down during that time.
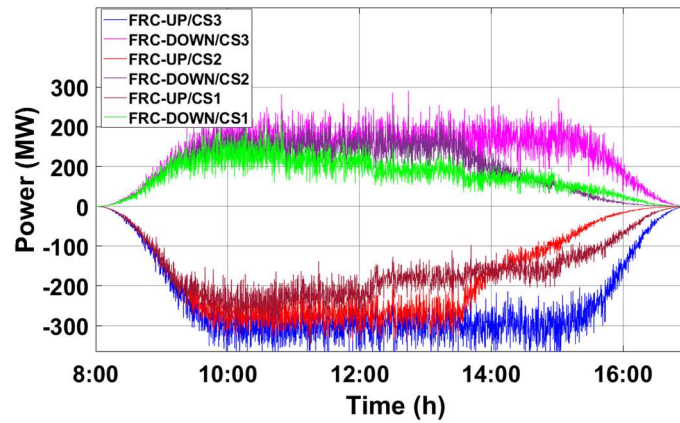

Fig. 15 EVs' FRC of regulation-up and regulation-down under CS1, CS2 and CS3.

In addition to achieving the charging demand of EVs using CS3, it is also able to keep the capacity of EVs to participate in frequency regulation and that is because the fast adjustments in the SCP of EVs resulted in the best suppression of ACE, frequency deviations and generators units' power as shown in Figs. 16, 17 and 18 respectively.
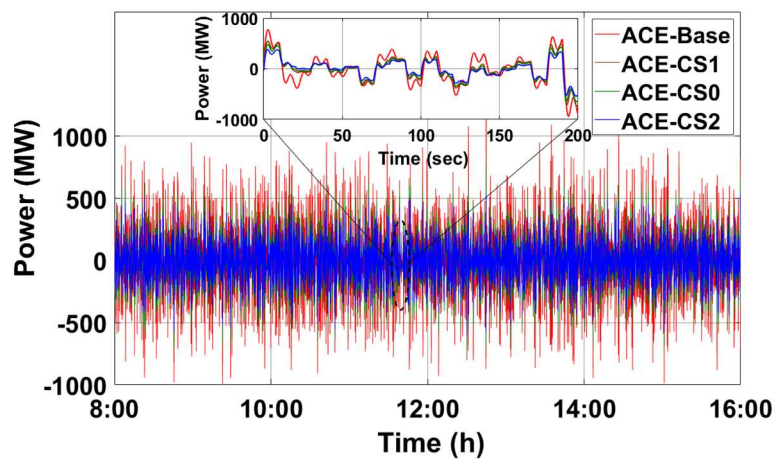

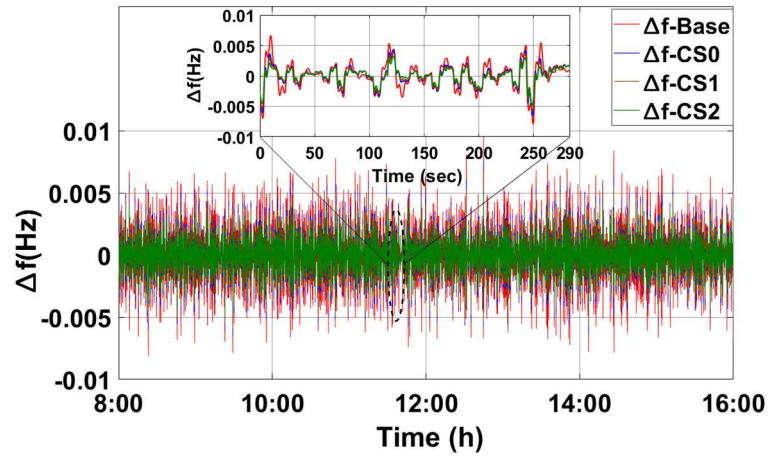Fig. 16 ACE signal after implementing CS0, CS1 and CS2.

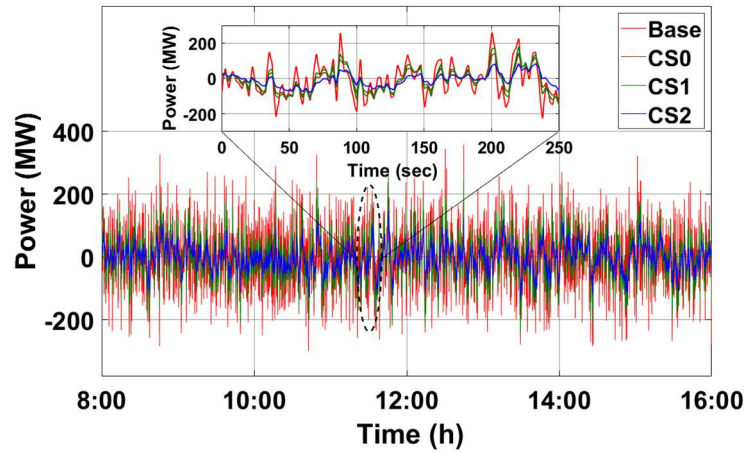Figure 17 System frequency deviation after implementing CS0, CS1 and CS2.

Fig. 18 Dispatch power allocated to generators units for frequency regulation.

While considering CS1, CS2, and CS3 using ACE and ARR, the grid frequency deviation and ACE quality in Area A are given in Tables 4 and 5 respectively. The Root Mean Square values (RMS) of the frequency deviations and ACE are calculated and respectively are also provided in Tables 4 and 5, to demonstrate the improvement achieved in the frequency regulation of Area A using the proposed V2G control strategy.

Table 4 Frequency Deviations in Area A.

Strategy	RMS of frequency deviation (Hz)	Percentage of decrease (%)
Basic Case	0.02784	--
CS1-ACE	0.02333	16.20
CS1-ARR	0.02284	17.96
CS2-ACE	0.02273	18.35
CS2-ARR	0.02165	22.23
CS3-ACE	0.02170	22.07
CS3-ARR	0.01961	29.57

Table 5 ACE in Area A.

Strategy	RMS of ACE (MW)	Percentage of decrease (%)
Basic Case	129.52	--
CS1-ACE	107.81	16.76
CS1-ARR	106.18	18.02
CS2-ACE	101.09	21.95
CS2-ARR	96.85	25.22
CS3-ACE	99.54	23.15
CS3-ARR	88.57	31.62

Although all strategies (CS1, CS2 and CS3) were able to reduce the frequency deviation and ACE of the grid, there are some differences between them. Firstly, CS1-ACE and CS1-ARR strategies have the lowest RMS of frequency deviation (16.20% and 17.96%, respectively) and lowest RMS of ACE (16.76% and 18.02%, respectively). This is because when CS1 strategy is used, the gradual adjusting in SCP causes a reduction in the regulation capacity of regulation-up or regulation-down throughout the plug-in period. In addition, this strategy exhibits poor performance in achieving the expected charging demand as discussed in Section 4.1.

CS2 has higher percentages of reduction in the RMS of the frequency deviation (18.35% and 22.23%) and RMS of ACE (21.95% and 25.22%) using CS2-ACE and CS2-ARR. It is also able to satisfy the driving demand of EVs' owners. However, when compared to CS3, it can be noted that CS3-ACE and CS3-ARR have the highest percentages of reduction in the RMS of the frequency deviation (22.07% and 29.57%) and RMS of ACE (23.15% and 31.62%), besides their ability to optimally achieve the expected SOC levels of EVs. This is because using the proposed strategy, the SCP is tracking and regulating in real time according to the frequency regulation task and the expected SOC of the EV battery. It is worth noting that the use of ARR signal in all strategies can ensure better frequency and ACE quality as compared to ACE dispatch.

C. Impacts of different SOC distribution on regulation

As presented in Table 3, a normal distribution under a variance of 0.01 is used to simulate the initial SOC of the EV battery. They may also have a higher variance. Therefore, to demonstrate the impacts of different initial SOC of EV batteries on the frequency regulation, both normal and uniform distributions with the same mean value are considered as shown in Fig.19. Although different initial SOC distributions are applied, it can be observed that the regulation is maintained at the same level and the expected charging demands are always satisfied for all three types of EVs.

4   Conclusion

This paper proposed an optimal Vehicle-to-Grid (V2G) control strategy for Supplementary Frequency Regulation (SFR) considering the regulation tasks received from the control centre and the expected charging demands of Electric Vehicles (EVs). Deep reinforcement learning is used for adjusting in real-time the scheduled charging power of the EV to satisfy the charging demand of the battery at plug-out time while performing frequency regulation. A Deep Deterministic Policy Gradient (DDPG) agent was used to automatically provide very fast decisions without a need for extensive calculations which has reduced the computational cost. The advantages of using DDPG agent result in significantly improving the SFR of EVs and satisfying the preferences of EVs' owners.
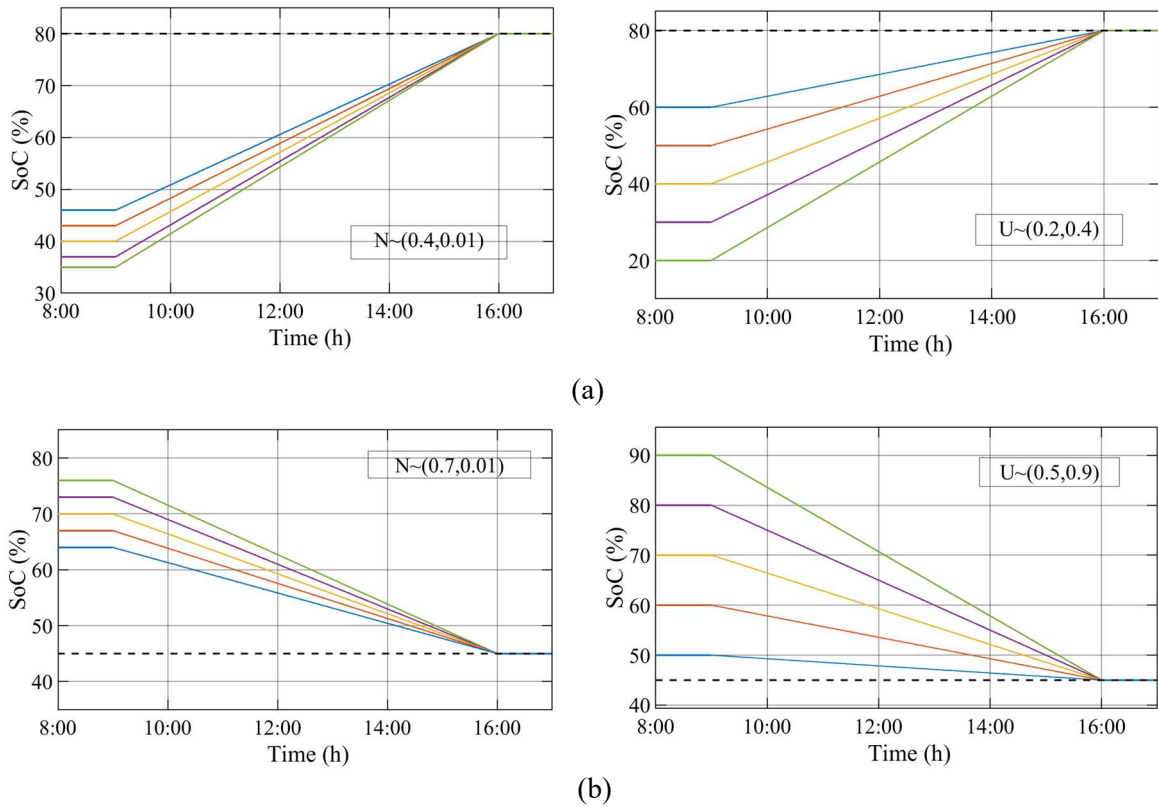
Fig. 19 SOC curves for (a) Type I and (b) Type II with normal and uniform distributions.

The proposed dispatch strategy is validated on a two-area interconnected power system. In this study, the EV aggregator was assumed to manage 100 EV charging stations each accommodating 500 EVs. The EV parameters such as arriving time, departure time, and the initial and the expected SOC of EV batteries are generated randomly using Monte Carlo method. The simulation results show an improvement in the quality of the frequency regulation while satisfying the charging demands of EVs. The proposed V2G strategy was shown to give a better performance in achieving the expected charging demand for the EVs participating in SFR as compared to other strategies even under large fluctuations in the Area Control Error (ACE) and large negative means values in the Area Regulation Requirement (ARR) signals.

5   References

[1] Shair, J., Li, H., Hu, J. and Xie, X., 2021. Power system stability issues, classifications and research prospects in the context of high-penetration of renewables and power electronics. Renewable and Sustainable Energy Reviews, 145, p.111111.

[2] Fernández-Guillamón, A., Gómez-Lázaro, E., Muljadi, E. and Molina-García, Á., 2019. Power systems with high renewable energy sources: A review of inertia and frequency control strategies over time. Renewable and Sustainable Energy Reviews, 115, p.109369.

[3] Tan, J. and Zhang, Y., 2017. Coordinated control strategy of a battery energy storage system to support a wind power plant providing multi-timescale frequency ancillary services. IEEE Transactions on Sustainable Energy, 8(3), pp.1140-1153.

[4] Li, J., Xiong, R., Yang, Q., Liang, F., Zhang, M. and Yuan, W., 2017. Design/test of a hybrid energy storage system for primary frequency control using a dynamic droop method in an isolated microgrid power system. Applied Energy, 201, pp.257-269.

[5] Meng, J., Mu, Y., Jia, H., Wu, J., Yu, X. and Qu, B., 2016. Dynamic frequency response from electric vehicles considering travelling behavior in the Great Britain power system. Applied energy, 162, pp.966-979.

[6] Datta, U., Saiprasad, N., Kalam, A., Shi, J. and Zayegh, A., 2019. A price-regulated electric vehicle charge-discharge strategy for G2V, V2H, and V2G. International Journal of Energy Research, 43(2), pp.1032-1042.

[7] Future Energy Scenarios (FES) Report. Published by UK's National Grid Operator, Warwick, UK (2018).

[8] Tan, K.M., Ramachandaramurthy, V.K. and Yong, J.Y., 2016. Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques. Renewable and Sustainable Energy Reviews, 53, pp.720-732.

[9] Habib, S., Kamran, M. and Rashid, U., 2015. Impact analysis of vehicle-to-grid technology and charging strategies of electric vehicles on distribution networks–a review. Journal of Power Sources, 277, pp.205-214.

[10] Vehicle to Grid Britain (V2GB) Report. Published by UK's National Grid ESO, UK (2019).

[11] Matt de Prez: New Fleets. [Online] (2020). https://www.fleetnews.co.uk/news/manufacturer-news/2020/05/21/fca-begins-v2g-pilot-at-mirafiori-plant.

[12] Almeida, P.R., Soares, F.J. and Lopes, J.P., 2015. Electric vehicles contribution for frequency control with inertial emulation. Electric Power Systems Research, 127, pp.141-150.

[13] Khooban, M.H., 2017. Secondary load frequency control of time-delay stand-alone microgrids with electric vehicles. IEEE Transactions on Industrial Electronics, 65(9), pp.7416-7422.

[14] Khalil, A., Rajab, Z., Alfergani, A. and Mohamed, O., 2017. The impact of the time delay on the load frequency control system in microgrid with plug-in-electric vehicles. Sustainable cities and society, 35, pp.365-377.

[15] Vita, V. and Koumides, P., 2019, September. Electric vehicles and distribution networks: Analysis on vehicle to grid and renewable energy sources integration. In 2019 11th Electrical Engineering Faculty Conference (BulEF) (pp. 1-4). IEEE.

[16] Meng, J., Mu, Y., Jia, H., Wu, J., Yu, X. and Qu, B., 2016. Dynamic frequency response from electric vehicles considering travelling behavior in the Great Britain power system. Applied energy, 162, pp.966-979.

[17] Liu, H., Huang, K., Yang, Y., Wei, H. and Ma, S., 2018. Real-time vehicle-to-grid control for frequency regulation with high frequency regulating signal. Protection and Control of Modern Power Systems, 3(1), pp.1-8.

[18] Wang, M., Mu, Y., Jia, H., Wu, J., Yu, X. and Qi, Y., 2017. Active power regulation for large-scale wind farms through an efficient power plant model of electric vehicles. Applied Energy, 185, pp.1673-1683.

[19] Amamra, S.A. and Marco, J., 2019. Vehicle-to-grid aggregator to support power grid and reduce electric vehicle charging cost. IEEE Access, 7, pp.178528-178538.

[20] Arias, N.B., Hashemi, S., Andersen, P.B., Træholt, C. and Romero, R., 2020. Assessment of economic benefits for EV owners participating in the primary frequency regulation markets. International Journal of Electrical Power & Energy Systems, 120, p.105985.

[21] Robson, S., Alharbi, A.M., Gao, W., Khodaei, A. and Alsaidan, I., 2021, April. Economic viability assessment of repurposed EV batteries participating in frequency regulation and energy markets. In 2021 IEEE Green Technologies Conference (GreenTech) (pp. 424-429). IEEE.

[22] Han, S. and Han, S., 2013. Economic feasibility of V2G frequency regulation in consideration of battery wear. Energies, 6(2), pp.748-765.

[23] Lam, A.Y., Leung, K.C. and Li, V.O., 2015. Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism. IEEE transactions on smart grid, 7(1), pp.156-166.

[24] Peng, C., Zou, J., Lian, L. and Li, L., 2017. An optimal dispatching strategy for V2G aggregator participating in supplementary frequency regulation considering EV driving demand and aggregator's benefits. Applied energy, 190, pp.591-599.

[25] Liu, H., Qi, J., Wang, J., Li, P., Li, C. and Wei, H., 2016. EV dispatch control for supplementary frequency regulation considering the expectation of EV owners. IEEE Transactions on Smart Grid, 9(4), pp.3763-3772.

[26] Liu, H., Huang, K., Wang, N., Qi, J., Wu, Q., Ma, S. and Li, C., 2019. Optimal dispatch for participation of electric vehicles in frequency regulation based on area control error and area regulation requirement. Applied energy, 240, pp.46-55.

[27] Wang, M., Mu, Y., Shi, Q., Jia, H. and Li, F., 2020. Electric vehicle aggregator modeling and control for frequency regulation considering progressive state recovery. IEEE Transactions on Smart Grid, 11(5), pp.4176-4189.

[28] Wang, M., Mu, Y., Li, F., Jia, H., Li, X., Shi, Q. and Jiang, T., 2019. State space model of aggregated electric vehicles for frequency regulation. IEEE Transactions on Smart Grid, 11(2), pp.981-994.

[29] Zhang, W., Gandhi, O., Quan, H., Rodríguez-Gallegos, C.D. and Srinivasan, D., 2018. A multi-agent based integrated volt-var optimization engine for fast vehicle-to-grid reactive power dispatch and electric vehicle coordination. Applied Energy, 229, pp.96-110.

[30] Ko, H., Pack, S. and Leung, V.C., 2018. Mobility-aware vehicle-to-grid control algorithm in microgrids. IEEE Transactions on Intelligent Transportation Systems, 19(7), pp.2165-2174.

[31] Arulkumaran, K., Deisenroth, M.P., Brundage, M. and Bharath, A.A., 2017. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), pp.26-38.

[32] Liu, H., Huang, K., Yang, Y., Wei, H. and Ma, S., 2018. Real-time vehicle-to-grid control for frequency regulation with high frequency regulating signal. Protection and Control of Modern Power Systems, 3(1), pp.1-8.

[33] Liu, H., Qi, J., Wang, J., Li, P., Li, C. and Wei, H., 2019. EV dispatch control for supplementary frequency regulation considering the expectation of EV owners. IEEE Transactions on Smart Grid, 9(4), pp.3763-3772.