Vehicle to Grid Frequency Regulation Capacity Optimal Scheduling for Battery Swapping Station Using Deep Q-Network

Xinan Wang, *Student Member, IEEE*, Jianhui Wang, *Senior Member, IEEE*,
and Jianzhe Liu, *Member, IEEE*

***Abstract*—Battery swapping stations (BSSs) are ideal candidates for fast frequency regulation services (FFRS) due to their large battery stock capacity. In addition, BSSs can precharge batteries for customers and the batteries that are not in charging can provide a stable regulation capacity to the market. However, uncertainties, such as ACE signals and the EV per-hour visit counts, introduce stochastic nonlinear dynamics into the operation of a BSS-based FFRS. Currently, there is no quantification method to ensure its optimal economical operation. To close this gap, in this article, we propose a novel deep Q-learning-based FFRS capacity dynamic scheduling strategy. This method can autonomously schedule the hourly regulation capacity in real time to maximize the BSS's revenue for providing FFRS. Case studies using real-world data verify the efficacy of the proposed work.**

***Index Terms*—Battery swapping station, deep Q-Network, fast frequency regulation service, vehicle to grid (V2G) service.**

NOMENCLATURE

Indices

n	Index referring to EV.
t	Index referring to time horizon by hour.
j	Index referring to regulation ACE signal.

Parameters

N	Hourly BSS visit count.
N_{bs}	Battery stock capacity of a BSS.
S_{ip}	Initial SOC in the battery to be precharged.
S_{ic}	Initial SOC in the coming EV.
$\bar{S}_f/\underline{S}_f$	Battery SOC upper/lower bound for participating in FFRS.
F_B	Lithium-ion battery price in \$/kWh
$\underline{P}_r/\bar{P}_r$	Per battery regulation capacity lower/upper limits.
F_{ch}	Charging price paid by EV owners (\$/kWh).
F_l	Locational marginal price.
$F_{p/c}$	FFRS performance/capacity clearing price.
NCY	Number of charge/discharge cycle per year
β	Ratio of BSS visit count to traffic flow.
η_b	Overall system efficiency for battery charge/discharge.
U	Battery capacity of an EV (kWh).
φ	Performance score.
$\delta^{+/-}$	Fractional regulation up/down signal.
μ_1	Statistical means q.
σ_1	Statistical standard deviations for q.
ϵ	Battery value depreciation rate by year

Variables

P_r	Per battery scheduled regulation capacity (kW).
q	Fractional hourly cumulative battery SOC gain due to FFRS.
λ	Mileage ratio.
R_t	Total scheduled FFRS capacity from a BSS.
S_f	SOC in a battery participating in FFRS.
$S_{io/ig}$	Battery's SOC that is lower/higher than the bounds $[\underline{S}_f, \bar{S}_f]$ due to participating in FFRS.
N_{pc}	Number of precharged batteries.
N_{pf}	Number of battery participating in FFRS.
N_{fc}	Number of batteries with SOC>$\bar{S}_f$ due to FFRS.
N_{od}	Number of batteries with SOC<$\underline{S}_f$ due to FFRS.
P_c	Energy needed to compensate the batteries with SOC< $\underline{S}_f$ to 50% SOC.
P_p	Energy needed to precharge batteries.
P_f	Energy needed to fully charge the batteries with SOC>$\bar{S}_f$.
B_{ch}	Revenue for charging service.
B_{fr}	Revenue from FFRS.
C_B	Cost of battery degradation.

Abbreviations

EV	Electrical vehicle.
TF	Traffic flow.
ACE	Area control error.

Manuscript received February 6, 2020; revised March 25, 2020; accepted April 23, 2020. Date of publication May 11, 2020; date of current version November 18, 2020. The work of Xinan Wang and Jianhui Wang was supported by the U.S. Natural Science Foundation under Grant 1800716. The work of Jianzhe Liu was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under the Solar Energy Technologies Office Award Number 34230. Paper no. TII-20-0609. *(Corresponding author: Jianhui Wang.)*

Xinan Wang and Jianhui Wang are with the Department of Electrical and Computer Engineering, Southern Methodist University, Dallas, TX 75205 USA (e-mail: xinanw@smu.edu; jianhui@smu.edu).

Jianzhe Liu is with the Energy Systems Division, Argonne National Laboratory, Lemont, IL 60439 USA (e-mail: jianzhe.liu@anl.gov).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2020.2993858

SOC	State of charge.
FFRS	Fast frequency regulation service
BSS	Battery swapping station.
DQN	Deep Q-learning Network.
AI	Artificial intelligence.
NN	Neural network.

I. Introduction

A COMPREHENSIVE study conducted by the Renewable Energy Institute (REI) [1] shows that renewable energy penetration in the U.S. reached 18% of the total generation in 2017, and this number is continuously growing. The inherent uncertainty and intermittent nature of renewables bring hazards to the stability of grid frequency, and therefore attract extensive research and attention [2]–[5]. One solution is to expand the regulation reserves [6]. Conventional regulation resources such as hydropower plants, combustion turbines, and steam turbines lack fast ramping flexibility [7], and therefore cannot satisfy the needs of the smart grid. On the other hand, faster ramping units, such as battery storage and flywheel energy storage can handle these challenges [7]. The motivation of this article is to concentrate EV battery resources to provide affordable and stable FFRS for the smart grid while guaranteeing the optimal economic benefit to the service provider.

As the EV penetration rapidly increases, the idea of using EV batteries to offer regulation services is proved to be technically feasible [8], [9] and attracts intense attention in academia [10]–[24]. However, the implementation of V2G-based FFRS faces several major challenges. The first challenge is scalability. The frameworks proposed in [10] and [11] enable the EVs parking at a single parking lot to participate in the FFRS and achieve the optimal charging. However, the number of EVs parking at a single facility is limited and they can hardly provide the minimum FFRS capacity (mostly 1 MW) required by the utilities [14]. Another challenge is the uncertainties of EV behaviors and ACE signals. The current ancillary service market requires an FFRS participant to maintain a stable regulation capacity on an hourly basis, which requires the EV fleets to dynamically adjust the regulation capacity of each EV to compensate for the capacity changes due to EV departures/arrivals and battery SOC limits. The V2G FFRS framework proposed in [12] considers the random EV behaviors as a Markov process and uses a Markov model to predict the FFRS capacity. The effectiveness of this framework relies on the model prediction accuracy and the optimization result may not be satisfactory if the model fails to reflect the fact. A robust V2G FFRS framework in [13] handles the EV and ACE uncertainties through a real-time greedy-index dispatch policy. This policy assumes all the EV owners are fully responsive to the designed incentive which compensates the FFRS-induced delayed-charging and battery degradation. The same assumption is made in [15], in which the droop control is adopted to share regulation capacities among EVs in proportion to their available battery capacities under the designed price incentive. However, as many researchers suggest [16], [17], it may not be realistic to assume the EV owners are willing to obey the regulation or responsive to a specific price incentive. The other challenge is the communication delay. FFRS requires the participants to respond to the ACE signals within a few seconds, failure to follow the ACE signals will lead to a low performance payment. EV aggregators-based FFRS control strategies shown in [18] and [19] require complex communication networks, which support EV aggregator to EV aggregator, EV to EV, and EV to aggregator information exchanges. Regardless of the control complexity, the associated communication delay means this method can hardly guarantee a timely response to the ACE signals for FFRS.

To tackle these major challenges, we propose to use battery swapping stations (BSSs) to provide FFRS in this article. A BSS can provide stable, sufficient, and zero delay FFRS capacity. Because a BSS does not need to worry about the EV owners' expected SOCs, and the number of batteries stocked in the BSS is sufficient to meet the capacity limit for FFRS [20]–[24]. However, the BSS-based FFRS still faces the challenges from EV behavior and ACE signal uncertainties. Moreover, the revenue model for FFRS might also involve market uncertainties. For instance, PJM includes the mileage ratio into their FFRS revenue model, which is decided by the real-time grid operation status [25]. Currently, there is no comprehensive solution to tackle these challenges and ensure the optimal economics of the BSS-based FFRS model. The frameworks for plug-in EVs introduced in [10]–[15], and [19] are infeasible to implement in BSS. For instance, the methods in [10], [13], and [15] need to collect every EV's arrival/departure schedule and SOC expectation. However, it is not possible for a BSS to accurately estimate the EV activities for a long duration, and the optimization solution will not be correct without an accurate prediction model. Other frameworks in [11], [12], [14], and [19] consider a large number of EVs and model the EV behaviors using certain distributions, such as normal distribution, Poisson distribution, etc. However, for a BSS which only serves a limited number of EVs per day, the EV uncertainties still exist and cannot be ignored. In addition, all those frameworks fail to consider the ancillary service market uncertainty.

In recent years, deep reinforcement learning (DRL) approaches have been successfully adopted in EV optimal charging scheduling [26]–[28] not only because it can handle the nonconvex relations between the EVs and the electricity market and always guarantee a feasible solution but also because of its real-time decision-making ability under severe uncertainties. In this article, we leverage the advantages of DRL and develop a DQN-based AI agent for the BSS-based FFRS to tackle the involved uncertainties in the nonconvex model and perform the optimal regulation capacity real-time scheduling for a BSS. Under this context, the main contributions of this article are as follows.

1) To the best of our knowledge, this is the first article that attempts to formulate the BSS-based FFRS as a stochastic dynamic problem and uses DQN for automatic optimal control of a BSS.
2) This framework not only handles the uncertainties from the EV behaviors and ACE signals but also deals with the uncertainties of the ancillary service market.
3) The practicality of the case studies in this framework is guaranteed by using real-world traffic data, ACE signals, and FFRS market data.

The rest of this article is organized as follows. Section II introduces the BSS-based V2G FFRS model and the associated uncertainties. Section III presents the BSS's economic model and the DQN agent training process for the proposed framework. Section IV shows the case studies that use real-world data to verify the effectiveness and economic feasibility of the proposed model, the test results are presented. Finally, Section V concludes this article.

II. Capacity Scheduling Strategy and Uncertainties

A. Regulation Capacity Scheduling Strategy

Like a gas station, a BSS functions as a centralized energy distribution center that provides instant energy services to EV owners. It can exchange energy with the grid by battery charging and discharging. In this article, we assume every EV needs to submit a service request to a BSS in advance so that the BSS can precharge batteries for them. To ensure seamless services to EV owners, a number of batteries equal to the BSS hourly service requests are precharged. The remaining batteries in the BSS can participate in FFRS. Batteries with an SOC that is out of a predefined bound $[\underline{S}_f, \bar{S}_f]$ must quit the FFRS in the next hour because their ramping capacities are insufficient. Those batteries with SOC $> \bar{S}_f$ are then fully charged in the next hour and replace the ones in the visiting EVs. If the amount of fully charged batteries exceeds the number of visiting EVs, the excessive batteries will be held to serve the next hour's visiting EVs. Those batteries with SOC $< \underline{S}_f$ are charged to a 50% SOC in the next hour and then put them back again to provide FFRS because a 50% SOC provides a battery with the equal ramping up and down capacity scheduling potential. The replaced batteries from the EVs are put together with the battery stock in the BSS to participate in FFRS. To minimize the charging cost, the batteries selected to be precharged for visiting EVs are several of the highest SOC batteries in the stock. FFRS requires a scheduled unit to maintain a constant regulation capacity on an hourly basis [7]. So that each battery maintains a fixed regulation capacity within each hour, and in our model, every battery has the same regulation capacity within each hour. The Algorithm I is the pseudocode for this introduced regulation capacity scheduling model.

Such a regulation capacity scheduling strategy indicates that the hourly available number of batteries in a BSS to provide FFRS dynamically changes in accordance with the EV visit count and the FFRS service load, which results in an uncertain hourly available FFRS capacity. The EV visit count N is stochastic and uncontrollable by the BSS, but the FFRS service load can be managed by adjusting the battery's hourly regulation limit $P_{r,t}$. Therefore, the BSS's optimal economical operation can be regarded as a problem of stochastic dynamic programming, there is in need of a strategy to determine the optimal hourly regulation limit $P_{r,t}$ for batteries in a BSS to guarantee the BSS's optimal economic operation.

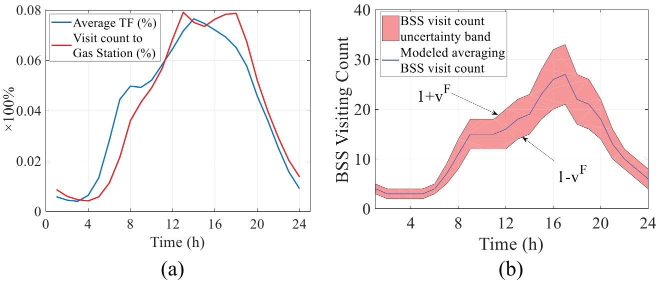
B. Uncertainty From BSS Visit Count

We assume that the customers' adoption of BSSs is the same as the customers' adoption of gas stations. Under such an assumption, we build a model that loosely binds the BSS's daily visiting profile within an uncertainty band. The model contains two stages. 1) Collect the historical hourly TF data $N_{f,t}$ for the place where the BSS is located; 2) converting the TF data into the BSS visit count N_t; the N_t is bounded by an uncertainty band ν^F. The validity of this model is justified based on the analysis shown below.

GasBuddy [30], examined more than 32.6 million consumer trips to gas stations and convenience stores around the U.S. in the first quarter of 2018, and they generate a gas station hourly visit percentage chart, and it plots as the red curve in Fig. 1(a). The blue curve is the daily average hourly TF percentage for 120 days' [29] TF count from a measuring station on road I-280 in San Jose, California, in 2017. The two curves nicely match with each other, and this match exists in the rest of the TF data we collect as well. Hence, we have our second assumption, i.e., the actual BSS visit count is positively linearly related to TF, as

Algorithm I: Regulation Capacity Scheduling Strategy.

```
Result: Determine the scheduling capacity $\hat{R}_t$
// set the initial battery SOCs in the BSS
Set $\boldsymbol{S}_{f,t} \rightarrow \{S_i | \underline{S}_f < S_i < \overline{S}_f\}$
while $t \leq T$ do // 24 hour per cycle, start from t=1
    $\boldsymbol{S}_{f,t-1} - \frac{\sum_{j=1}^{J}(\frac{\delta_{j,t}^{+}}{\eta_b} + \delta_{j,t}^{-} \cdot \eta_b) \cdot \Delta t}{Q} P_{r,t} \rightarrow \boldsymbol{S}_{f,t}$; // update the SOC of batteries participating in the FFRS on an hourly basis.
    $\{S_i | S_i \in \boldsymbol{S}_{f,t} \wedge S_i > \overline{S}_f\} \rightarrow \boldsymbol{S}_t^c$; // pick batteries with SOC higher than $\overline{S}_f$.
    $\{S_i | S_i \in \boldsymbol{S}_{f,t} \wedge S_i < \underline{S}_f\} \rightarrow \boldsymbol{S}_t^d$; // pick batteries with SOC lower than $\underline{S}_f$.
    $\boldsymbol{S}_{f,t} \setminus (\boldsymbol{S}_t^c \cup \boldsymbol{S}_t^d) \cup \boldsymbol{S}_{t-2}^d \cup \boldsymbol{S}_{t-1}^v \rightarrow \boldsymbol{S}_{f,t}$; // exclude $\boldsymbol{S}_t^c$ and $\boldsymbol{S}_t^d$ from $\boldsymbol{S}_{f,t}$; combine $\boldsymbol{S}_{t-2}^d$ and the batteries swapped from EVs $\boldsymbol{S}_{t-1}^v$ .
    if $|\boldsymbol{S}_{t-1}^c \cup \boldsymbol{S}_t^r| > |\boldsymbol{S}_{t+1}^v|$ then // "| |" denotes the number of elements in the set; $\boldsymbol{S}_t^r$ denotes the redundant fully charged batteries at hour t.
        $\emptyset \rightarrow \boldsymbol{S}_{t+1}^p$; // no need to pre-charge batteries
        $\boldsymbol{S}_{t-1}^c \cup \boldsymbol{S}_t^r \setminus \boldsymbol{S}_{t+1}^v \rightarrow \boldsymbol{S}_{t+1}^r$; // save the redundant batteries to $\boldsymbol{S}_{t+1}^r$.
    else
        $sort(\boldsymbol{S}_{f,t+1}, 'descent') \rightarrow \boldsymbol{S}_{f,t+1}$; // sort $\boldsymbol{S}_{f,t+1}$in descent manner.
        $\boldsymbol{S}_{f,t+1}^{(|\boldsymbol{S}_{t-1}^c \cup \boldsymbol{S}_{t-1}^r| - |\boldsymbol{S}_{t+1}^v|)} \rightarrow \boldsymbol{S}_{t+1}^p$; // select the top $|\boldsymbol{S}_{t-1}^c \cup \boldsymbol{S}_{t-1}^r| - |\boldsymbol{S}_{t+1}^v|$ SOC batteries from $\boldsymbol{S}_{f,t+1}$ to pre-charge for coming EVs.
        $\emptyset \rightarrow \boldsymbol{S}_{t+1}^r$; // no redundant batteries.
    end
    $\boldsymbol{S}_{f,t} \setminus \boldsymbol{S}_{t+1}^p \rightarrow \boldsymbol{S}_{f,t+1}$; // update the batteries participating in FFRS at time t+1.
    $P_{r,t+1} \cdot |\boldsymbol{S}_{f,t+1}| \rightarrow \hat{R}_{t+1}$; // get FFRS scheduling capacity $\hat{R}_{t+1}$ for t+1.
    set $t + 1 \rightarrow t$;
end
```



(a) (b)

Fig. 1. (a) TF count versus gas station visit. (b) BSS visit count estimation.

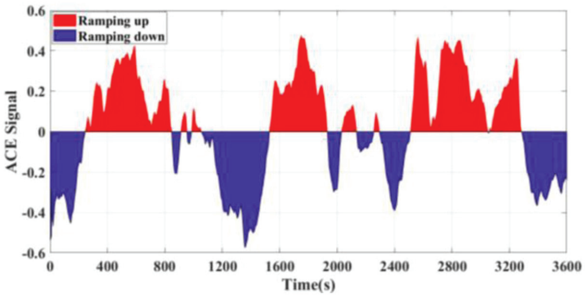


Fig. 2. ACE signal δ_t from 00:00 to 01:00 AM in 01/01/2017 from PJM.

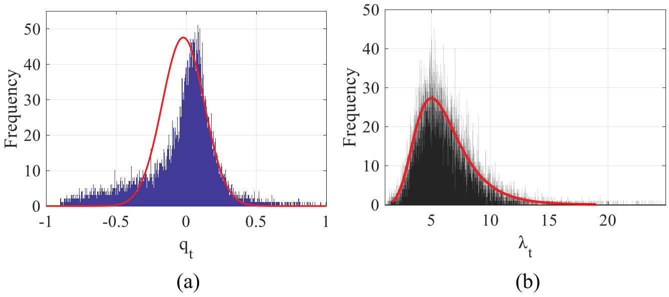


Fig. 3. (a) Distribution of q_t. (b) Distribution of λ_t.

follows:

$$N_t = \beta N_{f,t}, \quad t_s \le t \le t_e. \tag{1}$$

t_s and $t_e$are the start and end times of the interested period. β is the EV visiting ratio. To model the uncertainty of EV visit count, we apply an uncertainty band ν^F on the BSS visiting ratio β as shown in (2) and Fig. 1(b). The true 24-h EV visit count profile can be in an arbitrary shape as long as it is within the red area bounded by ν^F

$$\left(1 - \nu^F\right) \beta N_{f,t} \le N_t \le \left(1 + \nu^F\right) \beta N_{f,t}, \ \forall t. \tag{2}$$

C. Uncertainty From ACE Signals and Mileage Ratios

Batteries participating in FFRS are obligated to follow ACE signals $\boldsymbol{\delta} = \{\delta_t \in [-1, \ 1]\}$. The upper and lower bounds refer to the full scheduled capacities [31]. Fig. 2 shows the PJM ACE signal plot between 00:00 to 01:00 AM on 01/01/2017. The ACE signals bring severe uncertainties to the battery SOCs as shown in (3), where η_b is the battery charging/discharging efficiency, $\delta^+_{j,t}$ is the ramping up signal at time slot *j* in hour *t*, $\delta^-_{j,t}$ is the ramping down signal at time slot *j* in hour *t*. q_t decides the SOC change of each battery at hour *t* due to participation in FFRS. When q_t is negative, the battery SOC will increase; when it is positive, the battery SOC will decrease.

$$q_t = \sum_{j=1}^{J} \left(\frac{\delta^+_{j,t}}{\eta_b} + \delta^-_{j,t} \cdot \eta_b \right) \cdot \Delta t \tag{3}$$

$$q_t \sim \mathcal{N}\left(\mu_1, \sigma_1^2\right). \tag{4}$$

According to PJM 2017 and 2019 historical data, q_t's value $(\eta_b = 0.9)$ can be best fitted using a normal distribution as shown in Fig. 3(a) with its mean value -0.0216 and a standard deviation of 0.1508. It is seen that the mean value is on the left side of the peak value, this is because the left tail of the data is larger than the right tail, which pushes the mean value shifting to the left. We model the uncertainty of q_t using standard normal distribution $\mathcal{N}(-0.0216, 0.1508^2)$. When the SOC of a battery exceeds $\bar{S}_f$ or is under $\underline{S}_f$, that battery has to quit the next hour's FFRS and results in a decrease in the total FFRS available capacity of the BSS at the next hour. Therefore, the decision made on $P_{r,t}$ will impact the BSS's current and future income.

Mileage ratio, λ_t, is a market parameter that measures the relative work (movement) of fast ramping resources relative to conventional ramping resources, it plays an important role in the PJM's FFRS model [33]. Since the real-time mileage ratio depends on the grid operation status, we model it as a random variable in the dynamic control process. We collect the PJM 2017 and 2019 mileage ratio data, and fit the 0-99.6 percentile λ_t data into a generalized extreme value (GEV) distribution, as shown in Fig. 3(b), with shape parameter $k_\lambda = 0.0355$, scale parameter $\sigma_\lambda = 0.8713$, and location parameter $\mu_\lambda = 5.0572$. In the DQN agent training process, the λ_t is randomly generated following the GEV distribution: $\lambda_t \sim \text{GEV}(k_\lambda = 0.0355, \ \sigma_\lambda = 0.8713, \ \mu_\lambda = 5.0572)$.

III. Problem Formulation and Modeling

A. Mathematical Modeling of BSS Economic Benefits

The gross profit of the proposed BSS model contains two parts: 1) Battery charging services (5a)–(5g); 2) FFRS services (6a)–(6d)

$$P_{p,t} = \left(N_{pc,t} - \sum_{n=1}^{N_{pc,t}} S^n_{ip,t} \right) \cdot U \tag{5a}$$

$$\begin{aligned} N_{red,t} &= H\left[N_{fc,t-2} + N_{red,t-1} - N_t\right] \\ &\quad \cdot \left(N_{fc,t-2} + N_{red,t-1} - N_t\right) \end{aligned} \tag{5b}$$

$$\begin{aligned} N_{pc,t} &= H\left[N_{t+1} - N_{fc,t-1} - N_{red,t}\right] \\ &\quad \cdot \left(N_{t+1} - N_{fc,t-1} - N_{red,t}\right) \end{aligned} \tag{5c}$$

$$H[n] = \begin{cases} 0, & n < 0 \\ 1, & n \ge 0 \end{cases} \tag{5d}$$

$$P_{c,t} = \frac{1}{2} \left(N_{od,t-1} + N_t - \sum_{n=1}^{N_{od,t-1}} S^n_{io,t-1} - \sum_{n=1}^{N_t} S^n_{ic,t} \right) \cdot U \tag{5e}$$

$$P_{f,t} = \left(N_{fc,t-1} - \sum_{n=1}^{N_{fc,t-1}} S^n_{ig,t-1} \right) \cdot U, \ \ \forall t \tag{5f}$$

$$B_{ch,t} = \sum_{n=1}^{N_t} F_{ch,t} \cdot U \cdot \left(1 - S^n_{ic,t}\right) - F_{l,t} \cdot \left(P_{p,t} + P_{c,t} + P_{f,t}\right). \tag{5g}$$

Equation (5a) is the energy needed to precharge the batteries based on the BSS visit request. $N_{\text{red},t}$ is the number of the redundant fully charged batteries at the hour *t*. $N_{\text{red},t}$ exists when $N_{fc,t-2} + N_{\text{red},t-1} > N_t$, which is shown in (5b). In (5c), $N_{pc,t}$ is the number of batteries that needs to be precharged at time *t*. Equation (5d) is a Heaviside step function, whose value is zero for negative arguments and one for positive arguments. Equation (5e) is the energy needed to charge the batteries with SOC $< \underline{S}_f$ to 50% SOC before putting them back to FFRS at the next hour. $N_{od,t-1}$ is the number of batteries with SOC $< \underline{S}_f$ due to FFRS at time *t*-1. N_t indicates the number of visiting EVs at time *t*. Equation (5f) is the energy needed to fully charge the batteries with SOC higher than $\bar{S}_f$ due to FFRS at time *t*-1, those batteries replace the depleted batteries in the visiting vehicles. Equation (5g) shows the profits received from charging service, the first term is the customers' payment for charging service which is based on the difference between the left SOC in their batteries and the full SOC; the second term is the cost for BSS to purchase the charging energies from the grid

$$\begin{aligned} N_{pf,t} = N_{pf,t-1} - N_{pc,t} - N_{od,t-1} - N_{fc,\ t-1} \\ + N_{od,t-2} + N_{t-1} \end{aligned} \tag{6a}$$

$$R_t = N_{pf,t} \cdot P_{r,t} \tag{6b}$$

$$B_{fr,t} = R_t \cdot \varphi \cdot (\lambda_t F_{p,t} + F_{c,t}) \tag{6c}$$

$$S^v_{f,t} = S^n_{f,t-1} - \frac{q_t \cdot P_{r,t}}{U} (\underline{P}_r \le P_{r,t} \le \bar{P}_r) \tag{6d}$$

Equation (6a)–(6d) is the BSS's financial gain from participating in the FFRS. Equation (6a) is the number of batteries participating in FFRS at hour *t*. Equation (6b) is the scheduled FFRS capacity. Equation (6c) is the PJM model [25] for profit received for participating in FFRS. Equation (6d) is the SOC update for batteries participating in FFRS.

Battery degradation should also be considered in the cost of the operation since additional charge/discharge cycles will be added to the EV batteries in the BSS. Therefore, we incorporate a widely accepted depth of discharge (DOD) battery cycle life model [34] into our framework, as shown in (7)

$$T_{\text{life}} = \frac{n\text{Life}\,(\text{DOD})}{\text{NCY}} \tag{7}$$

in which T_{life} refers to the battery life in years, which is calculated by dividing the battery cycle life $n\text{Life}(\text{DOD})$ by the number of charge/discharge cycles per year (NCY). The battery cycle life $n\text{Life}$ is a nonlinear function of DOD as introduced in [34]. We flatten the battery's annual value depreciation cost to each charge/discharge cycle at that year to be the per-cycle degradation cost as shown in (8). The battery degradation cost is calculated for each battery swapped from EVs and the batteries quit FFRS due to low SOC

$$C_{B,t} = \sum_{n=1}^{N_{od,t}+N_t} \frac{U \cdot F_B \cdot (1-\epsilon)^{T_n} \cdot \epsilon}{NCY}, (0 \le T_n \le T_{life}) \tag{8}$$

In (8), F_B denotes the battery price in \$/kWh; ϵ indicates the yearly value depreciation rate of a battery. $U \cdot F_B \cdot (1-\epsilon)^{T_n}$ refers to the remaining value of the *n*th battery at its age T_n. The value depreciation of this battery at the current year is $U \cdot F_B \cdot (1-\epsilon)^{T_n} \cdot \epsilon$. Therefore, the per-cycle aging cost of this battery is evenly distributed to the NCY operation cycles in that year. In our model, the age of each battery is uniform randomly generated between 0 to T_{life}.

The daily gross profit of this model is the 24-h summation of (5g), (6c), and (8) shown as (9), which is also our objective function

$$obj:\ \max\ \boldsymbol{B} = \sum_{t=t_0}^{t_n} (B_{ch,t} + B_{fr,t} - C_{B,t}). \tag{9}$$

The whole problem is a stochastic dynamic programming problem, in which the value of parameters N_t, q_t, and $\hat{q}_t$ are uncertain. The decision to be made at each time step is $P_{r,t}$. Different parameter values and decision making at one step might change the remained solution trajectory of the whole problem. To solve this complicated nonconvex problem, we introduce a DQN agent to learn the optimal decision-making strategy at every time step. In our model, the inputs of the problem are the FFRS day ahead market prices, including $\boldsymbol{F}_p$,$\boldsymbol{F}_c$, and $\boldsymbol{F}_l$, and the stochastic parameters $\boldsymbol{N}, \boldsymbol{q},$ and λ. The output of the problem is the trained DQN network which can schedule $\boldsymbol{P}_r$ in the way of maximizing a BSS's daily operation profit $\boldsymbol{B}$. From 5(a)–(9), we know that $\boldsymbol{B} = \Phi(\boldsymbol{N}_{pf}, \boldsymbol{N}_{od}, \boldsymbol{N}_{fc}, \boldsymbol{N}_{pc}, \boldsymbol{N}_{red}, \boldsymbol{N},\ \boldsymbol{P}_r)$, Φ denotes a BSS's financial model, while$[\boldsymbol{N}_{pf}, \boldsymbol{N}_{od}, \boldsymbol{N}_{fc}, \boldsymbol{N}_{pc}, \boldsymbol{N}_{red}]$ is generated by the nonconvex function Algorithm I: $[\boldsymbol{N}_{pf}, \boldsymbol{N}_{od}, \boldsymbol{N}_{fc}, \boldsymbol{N}_{pc}, \boldsymbol{N}_{red}] = \Psi(\boldsymbol{N}, \boldsymbol{q},\ \hat{\boldsymbol{q}},\ \boldsymbol{P}_r)$ (Ψ denotes the Algorithm I). Therefore, in the training process N_t and q_t are stochastically generated following their distributions introduced in Section II-B and C. The initial SOCs for the swapped batteries are uniform random generated from [0, 0.2]. The DQN agent learns to make action a_t (the value of $P_{r,t}$) based on the state $s_t = [N_{pf,t}, N_{od,t}, N_{fc,t}, N_{pc,t}, N_{red,t}, N_t]$. The action space and state space are both discrete, the dimension of action space is $\frac{\bar{P}_r - \underline{P}_r}{\tau}$, where $\bar{P}_r$ and $\underline{P}_r$ are the upper and lower bounds for $P_{r,t}$, and τ is $P_{r,t}$'s value incremental step. The size of state space is the product of dimensions of each element in the state vector s_t. Such a large state space and relatively small action space combination makes the DQN an ideal solver [36] for the problem. The training environment for the DQN agent is the BSS's operating model, which includes the battery management strategy in Algorithm I and the revenue models from (1)–(9). Fig. 4 is a flow chart summarizing the DQN agent's training environment. In Fig. 4, the solid line indicates the battery management flow of the BSS, and the dashed line refers to the revenue flow along the battery flow path.

B. Form Deep Q-Network

In reinforcement learning, an agent performs actions in a specific environment, and the environment responds to the actions by generating a new state, at the same time the agent receives a reward depending on what state it is in and what will be the next state when it performs the action, this process is shown in Fig. 5. In this manner, the agent is trained to maximize the total reward along the whole decision trajectory.

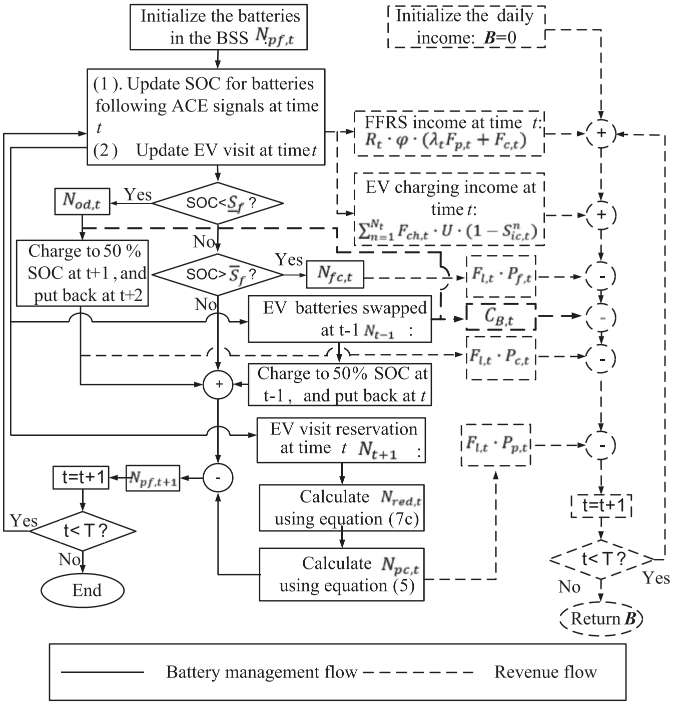

Initialize the batteries in the BSS $N_{pf,t}$
Initialize the daily income: $\boldsymbol{B}$=0
(1). Update SOC for batteries following ACE signals at time t
(2) Update EV visit at time t
FFRS income at time t: $R_t \cdot \varphi \cdot (\lambda_t F_{p,t} + F_{c,t})$
EV charging income at time t: $\sum_{n=1}^{N_t} F_{ch,t} \cdot U \cdot (1 - S_{ic,t}^n)$
$N_{od,t}$
Yes
SOC<$\underline{S}_f$?
No
Charge to 50 % SOC at t+1 , and put back at t+2
Yes
SOC> $\overline{S}_f$?
$N_{fc,t}$
No
EV batteries swapped at t-1 N_{t-1} :
$F_{l,t} \cdot P_{f,t}$
$C_{B,t}$
$F_{l,t} \cdot P_{c,t}$
+
Charge to 50% SOC at t-1 , and put back at t
EV visit reservation at time t N_{t+1} :
$F_{l,t} \cdot P_{p,t}$
t=t+1
$N_{pf,t+1}$
-
Calculate $N_{red,t}$ using equation (7c)
Calculate $N_{pc,t}$ using equation (5)
Yes
t< T ?
No
End
t=t+1
t< T ?
No
Yes
Return $\boldsymbol{B}$
Battery management flow
Revenue flow

Fig. 4. Flow chart for DQN agent's training environment.

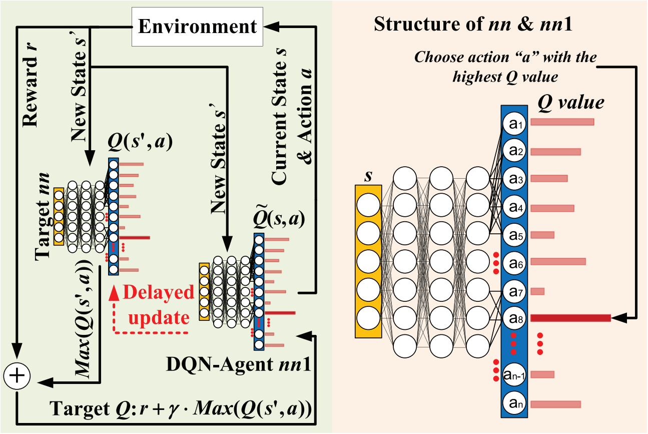


Fig. 5. DQN agent's training loop.

For a simple Q-learning, the agent learns the action-reward function $Q(s, a)$ in the manner of iteratively updating the Q value, as shown in (10), which is served to evaluate how good it is to take action a at state s. This equation is known as the Bellman equation which is also a necessary condition for optimality in dynamic programming. In (10), the term $Q(s_t, a_t)$ on the right side of the equation is the Q value of taking action a_t at state s_t based on the previous updated Q function; α denotes learning rate which discounts the Q updates to ensure the model doesn't overestimate the reward; r_t is the immediate reward at time t when action a_t is taken under the current state s_t; $\max Q(s_{t+1}, a)$ is the maximum possible Q value at the next state, this means the agent is looking forward to determining the best action to be taken to get the maximum future reward; γ is the discount factor which decreases the impact of future rewards impact on the current action decision-making. In our application, Q value is defined as the summation of the current operation profit B_t and anticipated discounted future reward B_{t+1}. If both the state space and the action space are small, the function Q can be formed into a Q-table to serve as a "cheat sheet" for the agent. However, if the action space and the state space are in thousands especially when states are in continuous form, it becomes inconvenient to learn and search in that huge table. In this context, an NN can be trained to interact with the environment and learn the sophisticated action-reward function Q. Then it can serve as an agent to take actions based on the current state

$$Q(s_t, a_t) = (1 - \alpha) Q(s_t, a_t) + \alpha \cdot (r_t + \gamma \cdot \max Q(s_{t+1}, a)). \tag{10}$$

To develop an NN which can perform Q-learning, its input should be the current state and some other information about the environment. In our setup, the state is $s_t = [N_{pf,t}, N_{od,t}, N_{fc,\ t}, N_{pc,t},\ N_t]$; the actions a_t the agent can take is the per-battery FFRS capacity limit $P_{r,t}$, the output is the Q value following the updating rule in (10). The reward function is the hourly profit B_t. The loss function of the NN is (11), which minimizes the difference between the predicted Q value $\tilde{Q}(s_t, a_t)$ given by the learning NN agent and the desired Q value $\hat{Q} = [r_t + \gamma \cdot \max Q(s_{t+1}, a)]$ based on the current reward r_t and discounted future reward estimated by a target NN. Notice that $\tilde{Q}(s_t, a_t) \neq Q(s_t, a_t)$, $\tilde{Q}(s_t, a_t)$ is given by a NN which updates at every step, $Q(s_t, a_t)$ is given by a target NN which has a delayed update, the reason will be introduced later

$$\mathcal{L} = \left\| r_t + \gamma \cdot \max Q(s_{t+1}, a) - \tilde{Q}(s_t, a_t) \right\|. \tag{11}$$

Algorithm II shows the full training process of a DQN agent in our framework. Some steps need to be explained in detail.

1) For each training episode, the daily EV visit count vector is generated stochastically following the distribution shown in (2).
2) ε-greedy action selection policy is implemented to avoid the training process to be locked in a locally optimal solution. Given the random nature of the environment, if the agent makes a wrong decision at the beginning, then this decision will continue to be made by the agent because it only selects the maximum Q in any state. However, the ε-greedy action selection policy allows the agent to jump out of the locked solution and randomly explore another action because of its conditional selection mechanism shown in the **If** function in Algorithm II.
3) Two neural networks are needed instead of one;these two neural networks have an identical structure, one neural network ($nn1$) updates at every training step, the other neural network (nn) serves as a target network that provides a target $\boldsymbol{Q}$ vector for $nn1$. The target network nn has a delayed update, in our application nn is updated every 24 steps, because if nn is also updating at every step or only use $nn1$, then the $nn1$'s training process is to minimize the difference between itself and a moving target [35]. This

```
Algorithm II: DQN Training.
Input: Day ahead FFRS performance/capacity price $\boldsymbol{F}_{p/c}$, locational marginal price $\boldsymbol{F}_{l}$.
Output: DQN for $P_{r,t}$ decision making.
Initialize $\alpha, \gamma, \varepsilon, \eta$ // set training parameter
Initialize the experience replay buffer $\boldsymbol{M}$.
$\boldsymbol{P}_r \leftarrow [\underline{P_r}: \mu: \overline{P}_r]$ // define an action pool
nn.initial // initialize the target neural network
nn1← nn // copy target NN to have learning NN
For i in range(number of episode)
    Initialize $\boldsymbol{N}_t$ // generate a stochastic EV visiting count vector
    $\boldsymbol{s}$←reset.enviroment(); // reset initial environment state
    $\varepsilon \leftarrow \varepsilon \cdot \eta$;// update epsilon search criteria
    $r\_sum$←0;// reset the total reward to be zero
    $tik$←0;// reset the time
    nn← nn1; //target neural network updates every episode
    While $tik < 25$:
        If rand(1) < $\varepsilon$: // epsilon greedy searching
            $a \leftarrow \boldsymbol{P}_r$(randi($|\boldsymbol{P}_r|$)) //random select an action
        Else:
            $a \leftarrow \boldsymbol{P}_r$(argmax(nn.predict($\boldsymbol{s}$))) //select the action which yields the maximum predicted reward.
        End
        $\boldsymbol{s}', r \leftarrow execute.env(\boldsymbol{s}, a)$ // execute the action $a$ in state $\boldsymbol{s}$ in the environment and get the immediate reward $r$ and new state $\boldsymbol{s}'$.
        Store the transition ($\boldsymbol{s}, a, r, \boldsymbol{s}'$) into $\boldsymbol{M}$
        $\hat{Q} \leftarrow r + \gamma \cdot \max(\text{nn.predict}(\boldsymbol{s}'))$ // calculate the target $\hat{Q}$.
        $\boldsymbol{Q} = \text{nn.predict}(\boldsymbol{s})$ // get the current $Q$ value vector for each action at state $\boldsymbol{s}$ using target neural network nn.
        $\boldsymbol{Q}(\text{index}(a \text{ in } \boldsymbol{P}_r)) = \hat{Q}$ // updates the $Q$ vector
        Sample a batch of transitions $\boldsymbol{D}$ from $\boldsymbol{M}$
        nn1.fit([$\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{D}}$], [$\boldsymbol{Q}, \boldsymbol{Q}_{\boldsymbol{D}}$], epoch=1) // train the learning neural network nn1
        $\boldsymbol{s} \leftarrow \boldsymbol{s}'$ // update the state vector
        $r\_sum = r\_sum + r$ // updates the total reward
    End
    r_list.append(r_sum) // record the total reward for each episode.
End
```

will cause a severely unstable training process. A delayed update in nn can provide a stable target to $nn1$.

4) A replay buffer should be deployed to store the historical state transition. For each training epoch, a mini batch of historic data $\boldsymbol{D}$ should be sampled from the buffer $\boldsymbol{M}$. $\boldsymbol{D}$ is then combined with the latest transition ($\boldsymbol{s}_t$, a, r,$\boldsymbol{s'}_t$) to train the neural network $nn1$. This process is necessary because allowing the agent to learn from earlier memories can speed up the learning and break undesired temporal correlation. Besides, because DQN training is a circulation process between neural network and environment, it is vital to allow DQN to sample the past state transitions in each training episode so that it does not overfit to the most recent cases.

The DQN training work is conducted on GPU with model NVIDIA GTX 960M 2 GB memory. The computer used is equipped with Intel(R) Core(TM) i7-4720HQ processor with a clock rate of 3.60 GHz and 16 GB memory.

IV. Case Study

In the case study, we perform dynamic regulation capacity scheduling work in different scenarios. To demonstrate the practicality of our strategy, we use real-world data to set up the environment. The detailed system parameters are shown in Table I and described here: 1) 2017 traffic data at the intersection of SR-17 and I-280 in San Jose, California from PeMS [29] is used. However, this data is measured on the highway, we scaled the data 10 times down to represent the TF on a local street where a BSS can possibly be built at; 2) FFRS market data from PJM in 2017 and 2019 [32] is collected and used in case studies, each case uses one day's data which is randomly selected; 3) For a typical DQN, the action space should be in a discrete manner [37]. In our work, the action to be taken refers to the hourly per-battery regulation capacity limit $P_{r,t}$ which is bounded by the predefined upper limit 35 kW and lower limit 15 kW. To discrete the action, we space the selectable $P_{r,t}$ at 0.5 kW intervals: $P_{r,t} \in [15 : 0.5 : 35]$, therefore the dimension of action space is 41; 4) we assume the batteries in a regular EV without participating in FFRS have high use intensity and experience 365 charge/discharge cycles per year on average, and according to our simulation results those EV batteries that participate in the FFRS experience 1 time more charge/discharge cycles or 730 cycles per year. With 80% discharge depth, the battery life for those batteries participating in FFRS lasts 6 years.

TABLE I
System Parameter Setting

Para	Value	Para	Value
η_b	0.9	F_l	2017/19 PJM [32]
N_{bs}	80	F_p	2017/19 PJM [32]
U	30kWh	F_c	2017/19 PJM [32]
φ	0.98	λ	2017/19 PJM [32]
TF data	PeMS [29]	φ	2017/19 PJM [32]
β	0.05	$\delta^{+/-}$	2017/19 PJM [32]
Initial battery in BSS	60	P_r	15 kW-35 kW
$\overline{S}_f/\underline{S}_f$	0.2/0.8	F_{ch}	\$0.12/kWh
F_B	280 [38]	T_{life}	6 years [34]
ϵ	0.167 [34]	NCY	730 cycles/yr

A. Case I: FFRS Capacity Scheduling Without Uncertainties

Case I is to verify the applicability of a DQN to our problem setup. Therefore, the stochastic parameters N_t, λ_t, and q_t are set to be deterministic: $\boldsymbol{N} = \beta \boldsymbol{N}_f$, $\lambda = [5]$, and $\boldsymbol{q} = [-0.1503]$. In such a way, Case I becomes a dynamic programming problem. The training process converges within 2000 episodes; each episode refers to a 24-h period. For each episode, the 24-h cumulative financial gain is recorded and plotted in Fig. 6.

The reward converges after 1700 episodes of training. For the first 200 episodes, the reward variation is from \$1050.11 to \$1239.43, and the mean reward is \$1168.51. For the last 200 episodes, the reward variation is from \$1241.60 to \$1291.08, and the mean reward is \$1271.08. When the training converges, the reward variation is reduced by 73.86%; in addition, the mean reward is increased by 6.25%. In Fig. 7(a), the blue bar shows the $P_{r,t}$ 24-h scheduling decision made by the trained DQN agent; and the pink line shows the total regulation capacity of the BSS for 24 h. To demonstrate the performance of the DQN agent, we design a regular agent that determines a constant regulation

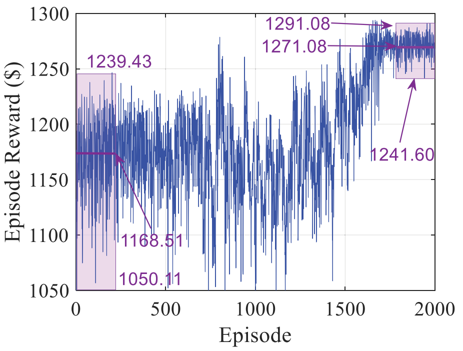


Fig. 6. DQN agent's episode reward during the training process.

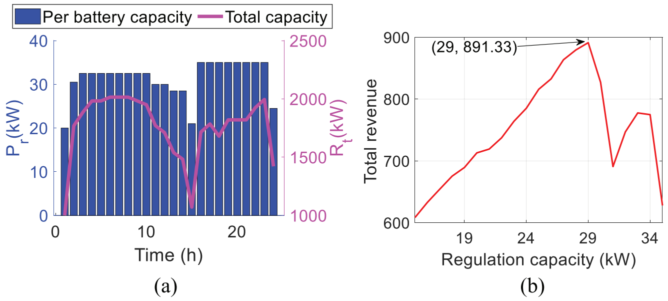


Fig. 7. (a) $P_{r,t}$ Scheduling comparison. (b) Reward for the regular agent.

capacity for batteries each day. Fig. 7(b) shows the reward received by the regular agent at different capacity levels, and the maximum reward received is \$891.33 when $\boldsymbol{P}_r = [29] kW$. According to Fig. 6, the DQN agent gains at least 39.30% more profit than the regular agent. This case verifies that a DQN agent can adapt to our designed action/state space and the environment.

B. Case II: FFRS Capacity Scheduling With Uncertainties From ACE Signals and Mileage Ratios

In Case II, we set $\boldsymbol{q}$ and λ as random variables following the distributions introduced in Section II: $q_t \sim \mathcal{N}(-0.0216, 0.1508^2)$, $\lambda_t \sim GEV(k_\lambda = 0.0355,\ \sigma_\lambda = 0.8713,\ \mu_\lambda = 5.0572)$. $\boldsymbol{N}$ is set as a deterministic parameter $\boldsymbol{N} = \beta \boldsymbol{N}_f$. Case II is to verify the DQN's applicability to scenarios that the BSS can decide the EV charging service load, such as bus fleets. Fig. 8 shows the reward for 2000 training episodes.

The reward converges after 1750 episodes. For the first 200 episodes, the reward variation is from \$1035.12 to \$1301.80, and the mean reward is \$1190.04. For the last 200 episodes, the reward variation is from \$1378.08 to \$1573.16, and the mean reward is \$1369.42. When the training converges, the reward variation is reduced by 26.85%; in addition, the mean reward is increased by 15.80%. In Fig. 9(a), the red bar shows the P_r 24-h scheduling decision made by the trained DQN agent, and the black line shows the total FFRS capacity scheduled from the BSS for 24 h. In the test case using real-world data, the trained DQN agent earns \$1399.47 for the day. While the maximum reward

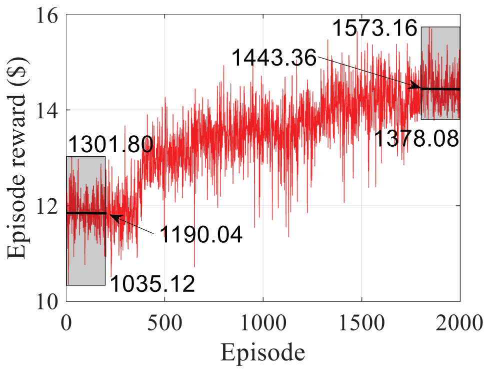


Fig. 8. DQN agent's episode reward during the training process.

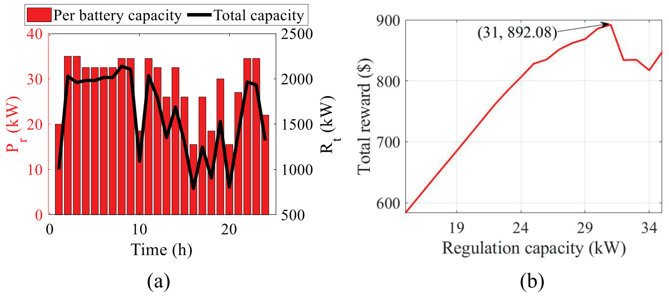


Fig. 9. (a) P_r Scheduling comparison. (b) Reward for the regular agent.

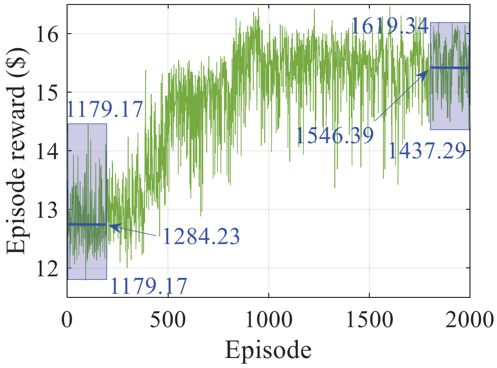


Fig. 10. DQN agent's episode reward during the training process.

received by the regular agent is \$892.08 when $\boldsymbol{P}_r = [31]$ kW as shown in Fig. 9(b). The DQN gains 56.88% more than the regular agent.

C. Case III: FFRS Capacity Scheduling With Uncertainties From ACE Signals, Mileage Ratios, and EV Visits

In Case III, we set $\boldsymbol{q}$, λ, and $\boldsymbol{N}$ as random variables; the values of $\boldsymbol{q}$, λ follow their distributions introduced earlier. For N_t, we set $\nu^F = 0.1$ to bound the N_t in the range $[0.9\beta N_{f,t}, 1.1\beta N_{f,t}]$. Fig. 10 shows the plot of reward for 2000 training episodes.

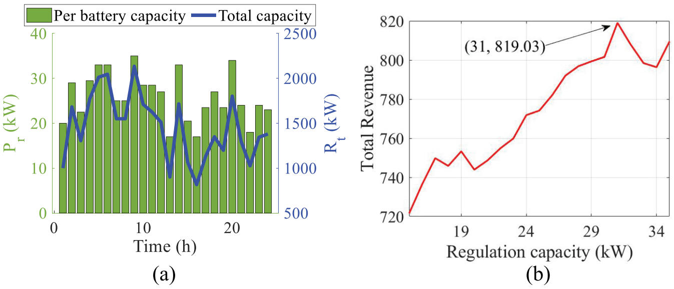


Fig. 11. (a) P_r Scheduling comparison. (b) Reward for the regular agent.

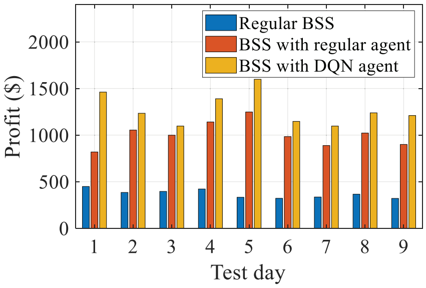


Fig. 12. 10-day profit comparison between three scenarios.

In this case, due to the large uncertainty involved in the environment, the episode reward does not converge as well as Case I and Case II. But the mean reward for the last 200 episodes still increases by 20.41% compared with the first 200 episodes. Fig. 11(a) shows the FFRS capacity scheduling decision made by both the DQN agent and the regular agent under real-world data. In this case, the regular agent gets the maximum reward of \$819.03 when $\boldsymbol{P}_r = [31]\ kW$. The DQN agent receives \$1461.29 which accounts for a 78.42% increase compared with the regular agent.

Case II and Case III show that the DQN can handle the large uncertainties in the environment and provide a satisfying FFRS capacity scheduling result for the BSS. However, the more uncertainties involved in the environment, the larger the reward variance will be. For a regular agent, the optimal FFRS scheduling capacity can be different every day; therefore, in practice, it is hard for a regular agent to make the optimal decision. In contrast, the DQN agent not only can adapt to a dynamic environment but also earns a much higher profit than the regular agent.

In Fig. 12, we show a 10-day profit comparison between a BSS participating in FFRS with a DQN agent, a BSS participating in FFRS with a regular agent, and a BSS without participating in FFRS. As discussed earlier, in this comparison these batteries in the regular BSS experience 365 charge/discharge cycles per year and their battery life lasts 12 years according to (8). These three models have the same service intensity and the same operation uncertainties. For the BSS with a regular agent, we assume the agent can make the optimal capacity decision $\boldsymbol{P}_r$ for every day. The results show that the BSS with a regular agent can make 2.72 times the profit of the BSS without participating in FFRS. The BSS with a DQN agent can make 3.45 times the profit of the BSS without participating in FFRS. The DQN agent can help a BSS to gain 26.72% more profit than the regular agent. The training time for each case is about 20–25 min, which makes it feasible to implement in the day-ahead market.

V. Conclusion

In this article, we proposed a comprehensive economic assessment model for a BSS to participate in FFRS. Because of the nonconvex nature and the stochastic parameters involved in the problem, we introduced the DQN agent to perform the optimal scheduling of the BSS regulation capacity. The results showed that a well-trained DQN agent can handle the large uncertainties in the model, and it was capable of making optimal dynamic decisions to ensure good profitability of the BSS. Although the profitability may vary for different system parameters and FFRS payment models, in the long run, as the battery cost drops and regulation demand increased, the BSS-based FFRS business model saw a promising future. Our case study was conducted based on real-world data, which makes the results very meaningful to the industry. The drawback of this method was that the BSS's available regulation capacity was determined on an hourly basis, therefore it could only passively participate in the ancillary service market as a price-taker. In the future, we will consider letting BSS participate in energy arbitrage so that the economic benefits of the BSS-based FFRS can be further enlarged by actively optimizing its charging/discharging activities. The associated bidding strategy will turn the problem into a more challenging multitime horizon optimization problem.

References

[1] R. Zissler, "Renewable energy in the US power sector, the other revolution," *Renew. Energy Inst.*, Tokyo, Japan, Jul. 2018.

[2] N. Nguyen and J. Mitra, "An analysis of the effects and dependency of wind power penetration on system frequency regulation," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 354–363, Jan. 2016.

[3] Y. Zhang, J. Wang, and Z. Li, "Uncertainty modeling of distributed energy resources: Techniques and challenges," *Current Sustain./Renew. Energy Rep.*, vol. 6, no. 2, pp. 42–51, Jun. 2019.

[4] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec. 2017.

[5] H. S. Jang, K. Y. Bae, H. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1255–1263, Jul. 2016.

[6] Q. Zhai, K. Meng, Z. Y. Dong, and J. Ma, "Modeling and analysis of lithium battery operations in spot and frequency regulation service markets in australia electricity market," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2576–2586, Oct. 2017.

[7] L. Thomas, "Energy storage in PJM exploring frequency regulation market transformation," *Kleinman Center Energy Policy*, Philadelphia, PA, 2017.

[8] A. Dutta and S. Debbarma, "Frequency regulation in deregulated market using vehicle-to-grid services in residential distribution network," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2812–2820, Sep. 2018.

[9] S. Izadkhast, P. Garcia-Gonzalez, P. Frìas, L. Ramìrez-Elizondo, and P. Bauer, "An aggregate model of plug-in electric vehicles including distribution network characteristics for primary frequency control," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2987–2998, Jul. 2016.

[10] G. R. C. Mouli, M. Kefayati, R. Baldick, and P. Bauer, "Integrated PV charging of EV fleet based on energy prices, V2G, and offer of reserves," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1313–1325, Mar. 2019.

[11] M. R. V. Moghadam, R. Zhang, and R. T. B. Ma, "Distributed frequency control via randomized response of electric vehicles in power grid," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 312–324, Jan. 2016.

[12] M. Wang *et al.,* "State space model of aggregated electric vehicles for frequency regulation," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 981–994, Mar. 2020.

[13] X. Ke, D. Wu, and N. Lu, "A real-time greedy-index dispatching policy for using PEVs to provide frequency regulation service," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 864–877, Jan. 2019.

[14] A. Y. S. Lam, K. Leung, and V. O. K. Li, "Capacity estimation for vehicle-to-grid frequency regulation services with smart charging mechanism," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 156–166, Jan. 2016.

[15] H. Liu, J. Qi, J. Wang, P. Li, C. Li, and H. Wei, "EV dispatch control for supplementary frequency regulation considering the expectation of EV owners," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3763–3772, Jul. 2018.

[16] K. Chaudhari, N. K. Kandasamy, A. Krishnan, A. Ukil and H. B. Gooi, "Agent-based aggregated behavior modeling for electric vehicle charging load," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 856–868, Feb. 2019.

[17] X. Wang, Y. Nie, and K. E. Cheng, "Distribution system planning considering stochastic EV penetration and V2G behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 149–158, Jan. 2020.

[18] K. S. Ko, S. Han, and D. K. Sung, "A new mileage payment for EV aggregators with varying delays in frequency regulation service," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2616–2624, Jul. 2018.

[19] K. Kaur, N. Kumar, and M. Singh, "Coordinated power control of electric vehicles for grid frequency support: MILP-based hierarchical control design," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3364–3373, May 2019.

[20] T. Zhang, X. Chen, Z. Yu, X. Zhu, and D. Shi, "A Monte Carlo simulation approach to evaluate service capacities of EV charging and battery swapping stations," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 3914–3923, Sep. 2018.

[21] P. Xie, Y. Li, L. Zhu, D. Shi, and X. Duan, "Supplementary automatic generation control using controllable energy storage in electric vehicle battery swapping stations," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 4, pp. 1107–1116, 2016.

[22] F. Li, H. Li, and D. Liu, "Fast frequency regulation of power system based on EV swap-charging station," in *Proc. IEEE Vehicle Power Propulsion Conf.*, Hangzhou, China, 2016, pp. 1–4.

[23] Y. Cheng and Z. Chengwei, "Configuration and operation combined optimization for EV battery swapping station considering PV consumption bundling," *Protection Control Modern Power Syst.*, vol. 2, Dec. 2017, Art. no. 26.

[24] X. Tan, G. Qu, B. Sun, N. Li,and D. H. K. Tsang, "Optimal scheduling of battery charging station serving electric vehicles based on battery swapping," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1372–1384, Mar. 2019.

[25] G. He, Q. Chen, C. Kang, P. Pinson, and Q. Xia, "Optimal bidding strategy of battery storage in power markets considering performance-based regulation and battery cycle life," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2359–2367, Sep. 2016.

[26] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019.

[27] N. Sadeghianpourhamami, J. Deleu, and C. Develder, "Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 203–214, Jan. 2020.

[28] T. Qian, C. Shao, X. Wang, and M. Shahidehpour, "Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1714–1723, Mar. 2020.

[29] *Caltrans Performance Measurement System (PeMS)*, California Department of Transportation, 2019. [Online]. Availabe: http://pems.dot.ca.gov/

[30] "Foot traffic report for the fuel & convenience retailing industry," GasBuddy, Minneapolis, MN, USA, 2018. [Accessed Jan 23, 2020]. [Online]. Avaliable: https://www.iab.com/wp-content/uploads/2018/05/GasBuddy-Foot-Traffic-Report-Q1-2018-1.pdf

[31] R. H. Byrne, R. J. Concepcion, and C. A. Silva-Monroy, "Estimating potential revenue from electrical energy storage in PJM," in *Proc. IEEE Power Energy Soc. General Meeting*, Boston, MA, USA, 2016, pp. 1–5.

[32] *PJM FFRS marketing data*, PJM Interconnection LLC, Norristown, PA, USA, Accessed: 2019. [Online]. Avaliable: http://www.pjm.com/markets-and-operations/data-dictionary.aspx

[33] S. Benner, "Performance, mileage and the mileage ratio," PJM Interconnection LLC, Norristown, PA, USA, Nov. 11, 2015. [Accessed Jan 15, 2020]. [Online]. Available: https://www.pjm.com/-/media/committees-groups/task-forces/rmistf/20151111/20151111-item-05-performance-based-regulation-concepts.ashx

[34] I. Duggal and B. Venkatesh, "Short-term scheduling of thermal generators and battery storage with depth of discharge-based cost model," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2110–2118, Jul. 2015.

[35] I. Durugkar and P. Stone, "TD learning with constrained gradients," in *Proc. Deep Reinforcement Learn. Symp.*, Long Beach, CA, USA, Dec. 2017, pp. 1–8.

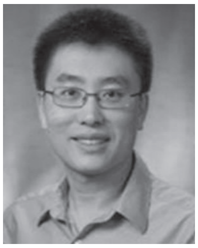
[36] R. Liu and J. Zou, "The effects of memory replay in reinforcement learning," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput.*, Allerton, IL, USA, 2018, pp. 478–485.

[37] D. Gabriel, R. Evans, P. Sunehag, and B. Coppin. "Reinforcement learning in large discrete action spaces," 2015, arXiv: abs/1512.07679.

[38] L. Goldie-Scot, "A behind the scenes take on lithium-ion battery prices," BloombergNEF, Mar. 5, 2019. [Accessed Jan 15, 2020], [Online]. Available: https://about.bnef.com/blog/behind-scenes-take-lithium-ion-battery-prices/

Xinan Wang (Student Member, IEEE) received the B.S. degree from Northwestern Polytechnical University, Xi'an, China, in 2013, and the M.S. degree from Arizona State University, Tempe, AZ, USA, in 2016, both in electrical engineering. He is currently working towards the Ph.D. degree in electrical and computer engineering at Southern Methodist University, Dallas, Texas, USA.

He was a Research Assistant with the AI & System Analytics Group at GEIRI North America, San Jose, CA, USA, in 2016, 2017, and 2019. His current research interests include machine learning applications to power systems, wide-area measurement systems, data analysis, and load modeling.



Jianhui Wang (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Illinois Institute of Technology, Chicago, Illinois, USA, in 2007. He is an Associate Professor with the Department of Electrical and Computer Engineering at Southern Methodist University, Dallas, TX, USA. He has authored and/or coauthored more than 300 journal and conference publications, which have been cited for more than 20 000 times by his peers with an H-index of 74. He has been invited to give tutorials and keynote speeches at major conferences including IEEE Conference on Innovative Smart Grid Technologies (ISGT), IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), IEEE Conference on Smart Energy Grid Engineering (SEGE), IEEE International Conference on High Performance and Smart Computing (HPSC) and International Green Energy Conference (IGEC-XI).

Dr. Wang was the recipient of the IEEE Power System Operation Committee Prize Paper Award (PES) in 2015 and the 2018 Premium Award for Best Paper in IET Cyber-Physical Systems: Theory & Applications. He was the 2018 and 2019 Clarivate Analytics highly cited researcher for production of multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science. He was an IEEE Power & Energy Society (PES) Distinguished Lecturer. He was the Editor-in-Chief for the IEEE Transactions on Smart Grid and is a Guest Editor for Proceedings of the IEEE special issue on power grid resilience.

Jianzhe Liu (Member, IEEE) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree in electrical and computer engineering from the Ohio State University, Columbus, OH, USA, in 2017.

He was a Visiting Scholar with the Aalborg University, Aalborg, Denmark, in 2017. He is currently a Postdoctoral Researcher at Argonne National Laboratory, Lemont, IL, USA. His current research interests include robust control and optimization for electric power systems.

Dr. Liu was the recipient of the 2019 Argonne Outstanding Postdoctoral Award.