Reinforcement Learning-Based Optimization of Capacity Configuration for Photovoltaic Energy Storage Charging Stations

Yang Liu, Mingda Song, Hengyu Liu, Zuoxia Xing, Qiwen Li, An Zhu

***Abstract*—The energy storage system in the photovoltaic energy storage charging station is one of the important components that ensure the stable operation of the system and increase the consumption rate of photovoltaic power generation. However, at present, photovoltaic energy storage charging stations are facing problems of unreasonable capacity configuration and high costs. To address this problem, this paper proposes a capacity configuration method for photovoltaic energy storage charging stations based on reinforcement learning. Firstly, by introducing a battery state of health model for the energy storage system, a corresponding configuration model is established to minimize the comprehensive cost of the photovoltaic energy storage charging stations under time-of-use electricity prices while considering constraints. Secondly, the Deep Q-Network algorithm is employed to solve the configuration model and obtain configuration results for photovoltaic energy storage charging stations, taking into account the battery's state of health. Finally, the effectiveness of the proposed model and method is validated through practical case analysis. The capacity configuration method for photovoltaic energy storage charging stations not only increased the economic benefits of the photovoltaic energy storage power stations by 12.66% but also provided technical references and theoretical support for the optimal design and configuration of photovoltaic energy storage power stations.**

***Index Terms*—photovoltaic energy storage charging station, energy storage system, capacity configuration, reinforcement learning**

I. Introduction

ELECTRIC vehicles (EVs) are a new type of transportation that uses electric energy to drive instead of traditional fuel-driven vehicles. With the rapid growth of EVs, many vehicles enter charging stations for charging, which has a significant impact on the distribution network where the charging stations are located. Therefore, the sharp increase in charging capacity exerts a notable influence on the entire power grid. The charging station serves as a critical supporting facility for EVs, with the Energy Storage System (ESS) being a core component of the Photovoltaic Energy Storage Charging Station (PESCS). Consequently, Therefore, it's necessary to rationally configure the capacity of the ESS. The optimal configuration of the ESS primarily entails planning its capacity, which determines the space-time translation capability of the system. However, the configuration of capacity is influenced by investment costs. Therefore, achieving an appropriate balance between ESS configuration and capital investment is crucial for realizing optimal ESS configuration [1-2]. To enable EVs to effectively utilize renewable energy sources, photovoltaic (PV) systems can serve as their primary energy source, thereby facilitating significant reductions in carbon emissions. Given the volatility of PV power generation, the reliability of power supply can be notably improved by configuring an ESS. Therefore, the configuration of an ESS within PESCS not only enhances grid stability but also allows new energy vehicles to harness renewable energy more efficiently [3-5]. During the planning phase for configuring an ESS, it is essential to thoroughly consider the potential loss of battery's state of health to optimize both the economic and environmental advantages. Given that time-of-use electricity pricing has been implemented in most regions, an effective energy management strategy should be developed based on time-of-use electricity pricing [6-8].

Currently, some progress has been made in the research on the capacity configuration of ESS, both domestically and internationally. Based on varying optimization objectives, the optimal configuration of ESS can be categorized into three distinct types: configurations aimed at economic efficiency, low carbon, and system reliability. In reference [9], a multi-agent deep reinforcement learning method is proposed for energy management of PESCS, which reduces the operating cost of PESCS. In reference [10], a synchronous capacity configuration and scheduling optimization model for integrated electric vehicle charging stations was proposed, which effectively improved economic benefits and reduced carbon emissions. In reference [11], V2G technology is integrated to construct a dual charging and discharging mode for EVs, a cost model of the charging system with maximum energy efficiency and minimum investment is established, and an optimal configuration scheme for the capacity of

Manuscript received March 11, 2025; revised August 1, 2025.

This work was supported by the Scientific Research Project of the Liaoning Education Department of China (LJ212410142059) and the Liaoning Province Science Technology Joint Plan (Fund) Project of China (2023-MSLH-263).

Yang Liu is an associate professor of the Shenyang University of Technology, Shenyang, Liaoning 110870, China (Corresponding author, e-mail: yangliu9611@163.com).

Mingda Song is a graduate student of Shenyang University of Technology, Shenyang, Liaoning 110870, China (e-mail: 597263817@qq.com).

Hengyu Liu is a doctoral student of Shenyang University of Technology, Shenyang, Liaoning 110870, China (e-mail: lhyat@126.com).

Zuoxia Xing is a professor of the Shenyang University of Technology, Shenyang, Liaoning 110870, China (e-mail: xingzuox@163.com).

Qiwen Li is a graduate student of Shenyang University of Technology, Shenyang, Liaoning 110870, China (e-mail: 2524030960@qq.com).

An Zhu is a graduate student of Shenyang University of Technology, Shenyang, Liaoning 110870, China (e-mail: 18941795250@163.com).

PESCS is proposed. In summary, conducting research on the optimal configuration of ESS capacity in the PESCS holds considerable significance. Whether the configuration of ESS is reasonable will directly affect the normal operation of PESCS and its overall economy [12-15].

In contrast, the deep Q-network (DQN) has outstanding advantages in achieving optimal operation for PESCS with uncertainties and complex nonlinear models. This is due to its two key characteristics: Dynamic relevance of decision-making: Reinforcement learning (RL) focuses on the current configuration problem and can make targeted decisions at different stages based on past experience and current status, aiming at maximizing long-term cumulative rewards. The ability to cope with complex systems: Under the architecture of PESCS, the types of equipment are becoming more and more abundant, and the degree of coupling between equipment is deepening, which brings great challenges to model optimization. Reinforcement learning can effectively deal with the situation of diverse and complex interactions of equipment through learning environment models and decision-making strategies, and then make decisions more accurately.

Consequently, this paper puts forward a capacity configuration approach to tackle the drawbacks in optimizing the ESS capacity configuration of the PESCS. By solving the ESS capacity configuration optimization model and taking into account the service life of ESS batteries, this approach ensures a reliable power supply and accomplishes peak load shifting. As a result, it enhances the overall operational efficiency of the PESCS. Eventually, the validity of the proposed model and methods is confirmed through case studies. The primary contributions of this research can be summarized as follows.

(I) In the capacity configuration of the PESCS, the impact of the ESS battery's state of health is taken into account. An optimal configuration model aiming to minimize the annual total cost of the PESCS is set up, and the effects of factors like the battery's state of health on the configuration outcomes are further examined.

(II) Using the time-of-use electricity price as a basis, developing a reasonable operation strategy are conducive to extending the battery's state-of-health. Then, a capacity configuration model for the ESS is established, taking into account the constraints related to the system power balance, ESS, and the cost associated with the battery's state of health.

(III) By constructing an optimal capacity configuration model that considers the health status of batteries and solving it using DQN algorithms, a capacity configuration method that can delay the battery's state of health and better adapt to energy complementary characteristics is obtained, which improves the economy of PESCS.

The structure of this article is organized as follows. In Section 2, the fundamental architecture and model of the PESCS are presented. Section 3 takes into account the battery's health status and formulates an optimized capacity-configuration model for the PESCS. Section 4 elaborates on the solution approach for the system's capacity configuration model. In Section 5, the feasibility of the proposed method is validated through a case analysis. Finally, Section 6 concludes the paper and offers perspectives on future research directions.

II. The Architecture Model of Photovoltaic Storage Charging Station

The fundamental structure of the PESCS is depicted in Fig. 1. It primarily consists of a PV system, an ESS, an energy management system for the PESCS, and an EV load section. Every component is linked to the DC bus and engages in energy exchange with the communication system via the energy management system [16].

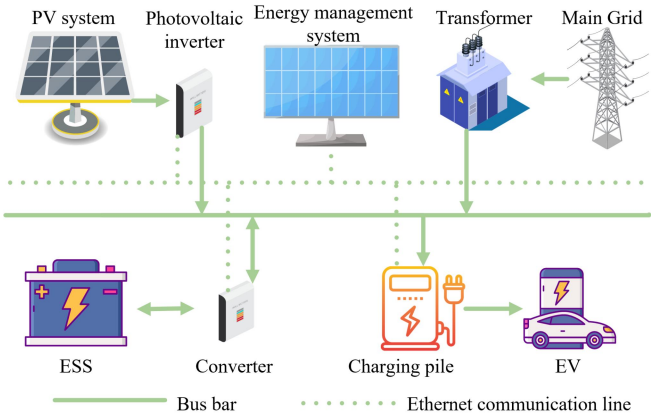

Fig.1. Basic structure of PESCS.

A. Photovoltaic generation model

The output power of the PV system is primarily influenced by factors such as light intensity and temperature. The output power $P_{PV}(t)$ at time t can be expressed as

$$P_{PV}(t) = \ln s(t) A \eta_c \eta_{pc} \left[1 + \nu \left(T_C(t) - T_{STC}\right)\right] \quad (1)$$

where, $\ln s(t)$ is the radiance (w/m^2), A is the area of the photovoltaic module, η_C is the photoelectric conversion efficiency of module, η_{PC} is the MPPT efficiency of the DC conversion link, ν is the weather correlation coefficient, $T_c(t)$ is the temperature of the PV cell in the t period, and T_{STC} is the temperature of the PV cell under standard test conditions.

B. Energy storage system model

The ESS at the charging station satisfies the EV charging demand via charging and discharging. The charging process model of the ESS is as follows.

$$SOC_{ESS}(t+1) = SOC_{ESS}(t) - \eta_{ESS} \frac{P_{ESS}(t) T_S}{C_{ESS}}, P_{ESS}(t) \leqslant 0 \quad (2)$$

The discharge process model of ESS is shown in the following equation.

$$SOC_{ESS}(t+1) = SOC_{ESS}(t) - \frac{1}{\eta_{ESS}} \frac{P_{ESS}(t) T_s}{C_{ESS}}, P_{ESS}(t) > 0 \quad (3)$$

where, $SOC_{ESS}(t+1)$ and $SOC_{ESS}(t)$ are the SOC of ESS at time $t+1$ and t, respectively; $P_{ESS}(t)$ is the power of ESS at time t; C_{ESS} and η_{ESS} are the capacity and charging efficiency of ESS, respectively; and T_S is the unit period.

C. EVs charging load model

The size of EVs random charging load is related to factors such as mileage, home time, and battery characteristics. Its

daily mileage m can be regarded as satisfying the logarithmic normal distribution, that is, $m \sim Log-N\left(\mu_{\mathrm{m}}, \delta_m^2\right)$. Its home time T_S can be regarded as satisfying the normal distribution, that is, $T_S \sim N\left(\mu_C, \delta_C^2\right)$. Then the probability density function satisfies the following formula.

$$f(m)=\frac{1}{m}\cdot\frac{1}{\sqrt{2\pi}\delta_m}\exp\left(-\frac{\left(\ln m-\mu_m\right)^2}{2\delta_m^2}\right) \tag{4}$$

$$f_c(T_S)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\delta_c}\exp\left(-\dfrac{\left(T_S+24-\mu_c\right)^2}{2\delta_C^2}\right) & 0<T_S\le(\mu_C-12)\\ \dfrac{1}{\sqrt{2\pi}\delta_c}\exp\left(-\dfrac{\left(T_S-\mu_C\right)^2}{2\delta_C^2}\right) & (\mu_C-12)<T_S\le 24\end{cases} \tag{5}$$

where, μ_m and δ_m are the average and standard deviation of EVs mileage, respectively; μ_C and δ_C are the mean and standard deviation of EVs at home.

The charging characteristic of EVs is usually to start charging after the last trip to the home, which is the starting time of EV charging. The charging process is simplified to a fixed power charge until it is fully charged. Then the charging duration of the ith EV is shown in the following equation.

$$T_i^{ch}=\frac{d_i\cdot\triangle E^{EV}}{\eta^{EV}\cdot P^{che}} \tag{6}$$

where, $\triangle E^{EV}$ is the power consumption per kilometer of EVs, P^{che} is the charging rated power of EVs, and η^{EV} is the charging efficiency of EVs.

The charging state of each EV in each period is determined by simulation. The cumulative random charging load of EVs in each time period can be determined by summing up the charging power of each individual EV during the corresponding period. The calculation formula is presented in Equation (7).

$$P_L=\sum_{i=1}^{N^{EV}}P^{che}\cdot A_{i,j}^{EV} \tag{7}$$

where, $A_{i,j}^{EV}$ is the judgment of the charging state of the ith EV at time t $1\le t\le 24$; $1\le i\le N$; P_L is the total random charging load of EVs at time t , and N^{EV} is the number of EVs.

D. System energy management strategy

First of all, battery state of health estimation can help to evaluate the impact on the battery's state of health and then choose the appropriate way to prolong the service life of the battery. Secondly, an effective operation strategy can improve the economy and reliability of the system. Therefore, there is a close relationship between a battery's state of health estimation and integrated energy configuration optimization. By accurately estimating the service life of the battery, a reasonable operation strategy can be formulated to achieve the optimization and efficient operation of the system. The energy management strategy of the PESCS is the basis for the energy exchange between the PESCS and the power grid and the basis for configuring the ESS capacity of the PESCS below [17-21]. The energy exchange strategy can be summarized as follows:

(I) When the PV system generates electricity, priority is given to the use of PV power.

(II) When the electricity price is low, in principle, the ESS is not used to charge the EVs, and the power supply is provided by the power grid. If the ESS is not fully charged, the ESS is charged by the power grid.

(III) When the electricity price is normal, the photovoltaic output is greater than the station load, and the state of charge of the ESS is not at full capacity. If the ESS cannot absorb all the PV power, the excess power is fed into the grid. If the PV output is less than the charging station load, the ESS is used to supply power to the station load. If the ESS cannot meet the load demand, the grid will cover the shortage.

(IV) During the peak period of electricity price, when the PV output is greater than the load and the state of charge of the ESS is not full, the excess PV power is absorbed by the ESS at this time. In the event that the ESS is unable to absorb the entirety of the PV power, the excess power is discarded. Conversely, when the PV output is lower than the load within the station, the ESS compensates for the power deficit.

For the convenience of calculation below, $P_{PV}(t)$ is the output power of the PV system at time t , $P_L(t)$ is the charging load at time t , $P_{ESS}(t)$ is the power of ESS at time t , and $P_G(t)$ is the energy exchange power between the PESCS and the power grid at time. When $P_{ESS}(t)<0$ represents the discharge of the ESS, and when $P_{ESS}(t)\ge 0$ represents the charging of the ESS. When $P_G(t)\ge 0$ indicates that the power grid supplies power to the PESCS, the PESCS purchases power from the power grid, and when $P_G(t)<0$ indicates the PESCS sells power to the power grid.

III. Capacity Configuration Model of Photovoltaic Storage Charging Station

A. Objective function

According to the PV power generation energy of the PESCS and the charging load of the EVs, the capacity configuration of the ESS of the PESCS is optimized. In the case of meeting the charging demand in the station, the design optimization goal is to minimize the annual total cost C of the PESCS.

$$\min \mathrm{C}=\min\left[\left(C_{gd}+C_{yw}+C_{ess,sh}+C_{grid}\right)-\left(C_{ch}+C_{bt}\right)\right] \tag{8}$$

where, C_{gd} is the initial fixed capital investment cost of the PESCS, C_{yw} is the operation and maintenance investment cost of the PV system and the ESS, $C_{ess,sh}$ represents the cost associated with the state of health of the ESS battery, C_{grid} is the purchase cost of the PESCS from the power grid when the PV system cannot meet the electricity demand, C_{ch} is the turnover obtained from the charging of the PESCS, and C_{bt} is the subsidy of the PESCS.

$$C_{gd}=P_{PV}C_{PV}\frac{r(1+r)^{n_1}}{(1+r)^{n_1}-1}+P_{ESS}C_{ESS}\frac{r(1+r)^{n_2}}{(1+r)^{n_2}-1} \tag{9}$$

where, P_{PV} , P_{ESS} is the rated output power of the photovoltaic array and the energy storage unit, C_{PV} , C_{ESS} is the unit price of the PV and ESS, n_1 is the Operational life of PV, and n_2 is the Operational life of ESS, r is the discount rate.

$$C_{yw}=\sum_{t=1}^{T}(K_{PV}P_{PV}C_{PV}+K_{PV}P_{ESS}C_{ESS})\Delta t \tag{10}$$

where, K_{PV}, K_{ESS} is the unit-rated maintenance cost of PV and ESS.

$$C_{ess,sh}=\frac{C_{ESS,initial}\left(1-SOH(t)\right)}{N_{rated}} \tag{11}$$

where, $C_{ESS,initial}$ is the initial cost of ESS, SOH_t is the health state of ESS, and N_{rated} is the number of battery quota cycles.

$$C_{gird}=\sum_{t=1}^{T}\left[\lambda_t\cdot P_{gird}\left(t\right)\right]\Delta t \tag{12}$$

where, λ_t is the price of electricity purchased by the PESCS from the superior power grid, and $P_{gird}\left(t\right)$ is the power that the PESCS needs to purchase from the superior power grid.

$$C_{ch}=\sum_{t=1}^{T}\left[\lambda_{ch}\left(t\right)P_L\left(t\right)\right]\Delta t \tag{13}$$

where, $P_L(t)$ is the EVs load of the PESCS, and $\lambda_{ch}\left(t\right)$ is the unit price of the power supply for the PESCS.

$$C_{bt}=\sum_{t=1}^{T}P_{PV}(t)\rho_{bt}\Delta t \tag{14}$$

where, ρ_{bt} is the state subsidy price of PV power generation per kilowatt-hour.

B. Constraint condition

The constraints of planning PESCS are mainly considered as the PV output system, charging and discharging of ESS, state of charge constraint, power supply reliability, and system power balance constraint.

Photovoltaic output constraints:

$$0\le P_{PV}\left(t\right)\le N_{PV}P_{PV}^{\max} \tag{15}$$

where, $P_{PV}^{\max}$ is the maximum output power of a unit PV.

Energy storage system state constraints:

The ESS should consider its own charge and discharge depth during each charge and discharge. The battery state constraint formula is as follows.

$$SOC_{ESS}^{\min}\leqslant SOC_{ESS}\left(t\right)\leqslant SOC_{ESS}^{\max} \tag{16}$$

where, $SOC_{ESS}^{\min}$ and $SOC_{ESS}^{\max}$ are the upper and lower limits of the state of charge of the ESS.

When the ESS is in the process of charging and discharging, taking into account factors like the lifespan of the ESS, the following is the constraint formula with its rated power as the maximum value.

$$0\le P_{ESS}\le P_{ESSout}^{max} \tag{17}$$

$$-P_{ESSout}^{max}\le P_{ESS}\le 0 \tag{18}$$

where,The maximum discharge power of the ESS is P_{ESSout}^{max}, and the maximum charging power of the ESS is represented is P_{ESSin}^{max}.

For battery management and battery's state of health prediction, SOH can reflect the health status and service life of the battery to a certain extent, so it plays an important role in battery management and battery's state of health prediction. In this paper, state of health (SOH) is used to estimate the health status of the battery. The battery discharge process expression is shown in the following equation.

$$SOH=\frac{C_i}{C_0}\times 100 \tag{19}$$

The cost associated with the state of health of the ESS battery during the i discharge process is denoted as C_{ess}^{i}, which can be defined as

$$C_{ess}^{i}=\frac{SOH_{i-1}-SOH_i}{80\%}C_d \tag{20}$$

where, C_i, C_0 and C_d represent the current full-charging capacity, nominal capacity, and initial cost of the ESS, respectively. When the SOH of the battery decays to approximately 20%, the battery is regarded as having reached the end of its service life.

System power balance constraints:

The power balance constraint is described by the chance constraint method. At any time t, EV charging load $P_L\left(t\right)$, charging station PV power $P_{PV}\left(t\right)$, ESS power $P_{ESS}\left(t\right)$, and power exchange with the grid $P_G\left(t\right)$ are in a balanced relationship.

$$\Pr\left(\sum_{i=1}^{N}P_L\left(t\right)=\sum_{t=1}^{T}P_{PV}\left(t\right)+\sum_{t=1}^{T}P_{ESS}\left(t\right)\pm P_G\left(t\right)\right)\ge\alpha_z \tag{21}$$

where, the probability of the event is $\Pr(\cdots)$, α_z is the confidence level at which the chance constraint holds.

IV. ENERGY STORAGE SYSTEM CAPACITY CONFIGURATION METHOD BASED ON REINFORCEMENT LEARNING

A. Modeling of Reinforcement learning

The basic components of RL include the state space s_t, action space a_t, and reward function $r\left(t\right)$, which represent the environment. According to the solving characteristics of RL, the capacity configuration optimization model of ESS in this paper is transformed into a DQN framework. Through multiple training of the deep reinforcement learning model, the optimal strategy is finally obtained to maximize the return of the entire scheduling cycle of the microgrid. Among them, state space, action space, and return function are the core elements of the whole process, which together constitute the deep reinforcement learning framework of microgrid optimal scheduling. According to the needs of the problem, the three elements are designed as follows. Its RL framework is composed of agents and an environment. The specific transformation process is as follows.

Establish state space:

In DQN, the state refers to the agent's perception of information from the external environment, and state space is a collection of environmental information. To avoid information redundancy and capture environmental information accurately and efficiently, an agent state perception model is established as

$$s_t=\left\{P_{PV}\left(t\right),P_L\left(t\right),P_G\left(t\right),SOC_{ESS}\left(t\right),SOH(t),\lambda_t\right\} \tag{22}$$

Action space:

Action is the action taken by the agent for the environmental state, and the charging and discharging action is shown in the following equation.

$$a_t\in\left\{-P_{ESS,MAX}^{c},\cdots,0,\cdots,P_{ESS,MAX}^{d}\right\} \tag{23}$$

where, $-P^{c}_{ESS,MAX}$ represents the maximum charging power, a negative value represents charging, $P^{d}_{ESS,MAX}$ represents the maximum discharge power, and a positive value represents discharge.

Reward function:

When the agent chooses an action under a state, the reward serves as the instant feedback it receives, which is the most important part of training the agent to achieve the goal.

The revenue reward is associated with the revenue that the PESCS acquires in the objective function, as presented in (24):

$$r_1(t)=\left(C_{gd}(t)+C_{yw}(t)+C_{ess,sh}(t)+C_{grid}(t)\right)-\left(C_{ch}(t)+C_{bt}(t)\right) \quad (24)$$

where, $r_1(t)$ is the revenue reward for the t period.

ESS health attenuation penalty: The ESS health decay penalty corresponds to the objective function battery's state of health cost C_{es}. After a capacity decay counting cycle T, the penalty factor α_K:

$$\alpha_k=\frac{E_b^{start}-E_b^{end}}{\sum_{t=1}^{T_1}\left|P_b(t)\right|}C_{es} \quad (25)$$

The immediate penalty for the capacity decay of the ESS as

$$r_2(t)=\alpha_k\left|p_b(t)\right| \quad (26)$$

The ESS SOH penalty is shown in (27).

$$r_3(t)=\lambda\left[1-\beta(x)\left(SOH(t)\geq 0.2\right)\right] \quad (27)$$

where, $\beta(x)$ is the index function; when x is positive, On the contrary, when the value of $\beta(x)$ is 1, the value of $\beta(x)$ is 0. When $SOH(t)<0.2$ is considered to be a deep discharge, $SOH(t)\geq 0.2$ is a cross-border penalty, it has a great influence on the battery's state of health. λ, which is a large number, serves as a penalty factor. The reward design takes into account behaviors that violate the SOH constraints.

In summary, the reward and punishment function of RL as

$$r(t)=\sigma_1 r_1(t)-\sigma_2 r_2(t)-\sigma_3 r_3(t) \quad (28)$$

where, σ_1、σ_2、σ_3 is the weight coefficient of the rewards and punishments of each part, and they are all positive.

State-action value function:

The DQN algorithm uses the state-action value function $Q^{\pi}(s,a)$, that is, the Q-value, to evaluate the long-term benefits of taking action a_t when the state s_t is taken, which can be expressed as

$$Q^{\pi}(s,a)=\mathbb{E}\left[\sum_{k=0}^{k}\gamma^k r_{t+k}\,\middle|\,s_t=s,a_t=a\right] \quad (29)$$

where, k is the range of time steps, and the value range is [0, 1], π is the strategy of mapping from the environmental state to the allocation strategy.

By relying on the Bellman equation, the state - action value function $Q^{\pi}(s_t,a_t)$ is capable of being formulated as

$$Q^{\pi}(s_t,a_t)=\mathbb{E}\left[r_t(s_t,a_t,s_{t+1})+\gamma Q^{\pi}(s_{t+1},a_{t+1})\right] \quad (30)$$

By iteratively updating the Q-value function, the Q-learning algorithm can gradually converge to the optimal Q-value function to find the optimal strategy. In the DQN algorithm, solving the optimal strategy π^* is equivalent to seeking the maximization of the state-action value function $Q^{\pi}(s_t,a_t)$:

$$\hat{Q}^{\pi^*}(s,a,\omega)\approx\max_{\pi}Q^{\pi}(s,a) \quad (31)$$

where, ω denotes the parameter of the neural network.

Consequently, the Bellman equation corresponding to the state-action value function $Q^{\pi}(s_t,a_t)$ can be formulated as

$$Q^{\pi}(s_t,a_t)=\mathbb{E}\left[r_t(s_t,a_t,s_{t+1})+\gamma\max_{\pi}Q^{\pi}(s_{t+1},a_{t+1})\right] \quad (32)$$

B. Solution based on an improved DQN algorithm

DQN is a value-based reinforcement learning algorithm. It defines the state action value function, namely the Q function, and iteratively learns by substituting the observation data into the Q function. Within the DQN algorithm, the state, action, and Q-value of each round are stored. The agent makes action recommendations by querying the Q-table. During the training phase of the agent, the Q-value is updated according to formula (33).

In the DQN algorithm, the state, action, and Q-value of each round are stored. The agent makes action recommendations by querying the Q-table. During the training of the agent, the Q-value as

$$\begin{aligned}&Q_{t+1}(s_t,a_t)=Q_t(s_t,a_t)+\\&\alpha\left[r_t+\gamma\max_{a_{t+1}}Q_t(s'_{t+1},a'_{t+1})-Q_t(s_t,a_t)\right]\end{aligned} \quad (33)$$

$$\alpha=\left(\frac{1}{2}(1+\cos(\xi_n\pi))(1-c_{\min})+c_{\min}\right)\alpha_0 \quad (34)$$

where, α and α_0 are the learning rate and initial learning rate, respectively, which are used to balance the importance of the agent to the current estimation and the previous accumulated experience; ξ_n is the cosine coefficient, and $C_{\min}$ is the minimum attenuation rate.

Therefore, the Bellman iteration equation based on the DQN architecture can be expressed as

$$\begin{aligned}&Q_{t+1}(s_t,a_t,\theta)=Q_t(s_t,a_t,\theta)+\\&\alpha\left[r_t+\gamma\max_{a_{t+1}}Q_{t+1}(s_{t+1},a_{t+1},\theta)-Q_t(s_t,a_t,\theta)\right]\nabla_{\theta}Q(s,a,\theta)\end{aligned} \quad (35)$$

Where, the network parameters of the evaluation network and the target network are θ^{+} and θ^{-}, respectively.

After accumulating a certain number of samples, the DQN algorithm extracts samples (s_j,a_j,r_j,s_{j+1}) from the experience replay unit for the loss function L_θ to update the evaluation network parameter θ and copies the parameter θ' to the target network at every N_f step.

$$L_\theta=\frac{1}{N_b}\sum_{j=1}^{N_b}\left[y_j-Q_{j+1}(s_{j+1},a_{j+1},\theta)\right]^2 \quad (36)$$

$$y_j=r_j+\gamma Q_{j+1}\left(s_{j+1},\arg\max_{a_{j+1}}\left(Q_{j+1}(s_{j+1},a_{j+1},\theta)\middle|\theta'\right)\right) \quad (37)$$

where, N_b is the mini-batch sample size, and y_i is the target Q value. By using two independent networks to estimate the action value function, DQN can be more accurate. Evaluate the value of the action and reduce the problem of

overestimation, thereby improving the performance of reinforcement learning.

C. DQN algorithm training process

Based on the Q-learning algorithm and a deep neural network, a DQN is developed. By using two independent networks to estimate the action value function, DQN can more accurately evaluate the value of the action and reduce the problem of overestimation, thereby improving the performance of RL. DQN improves the ability to process data, overcomes the shortcomings of traditional solvers that cannot process too much data, and has good practical significance.

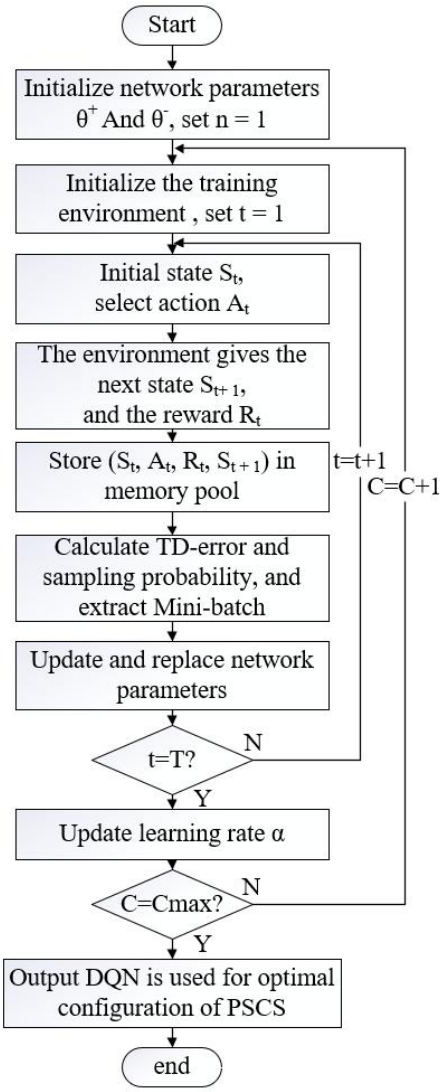

Fig. 2. PESCS capacity configuration training flow chart based on DQN.

Fig. 2 is the training flowchart of capacity configuration optimization of PESCS. First, initialize the network function and initialize the training environment. Secondly, after each action, the agent is selected to receive an instant reward. In each cycle, the accumulated learning experience is put into the memory pool, the TD-error of all samples in the sample pool is calculated, and the mini-batch is randomly extracted from the memory pool. Then, the gradient descent method is used to optimize the network parameters with the goal of minimizing the error loss function. After the training is completed, the weight parameters of the network are updated. After a certain training step, the network weights are copied to the target network to evaluate the reward values of all actions that can be taken in each state. At the end of each round, the learning rate of the agent is attenuated once to balance the exploration ability of the agent. Finally, the optimal parameter network is used to optimize the capacity configuration of the PESCS.

V. EXAMPLE ANALYSIS

A. Research Objects and Basic Data of PESCS

The basic simulation settings of the PESCS are shown in TABLE 1. Taking the PESCS shown in Fig. 1 as an example, the load of a typical solar storage charging station in Shenyang, Liaoning Province, is selected, and the configuration optimization of the ESS capacity of the PESCS is solved based on the RL algorithm. Taking to minimize the annual total cost of the PESCS as the optimization objective, the proposed configuration optimization model of the PESCS is simulated and verified.

TABLE 1
INTERRELATING PARAMETER

Related parameter names	Parameter value
PV service life (Year)	25
ESS service life (Year)	12
Unit ESS capacity cost (CNY/kWh)	1800
Unit PV installation cost (CNY/kW)	3000
Charging service fee (CNY/kWh)	1.8569
Operation maintenance coefficient	0.01
ESS charging/discharging efficiency	95%
Minimum/Maximum SOC of ESS	0.05-0.95
Discount rate	0.08
State subsidy for PV power generation (CNY/kWh)	0.08
Charging pile unit price (CNY)	20000

In terms of parameter setting of reinforcement learning, the number of hidden layers of the evaluation network and the target network is selected to be 3, the number of neurons in each layer is 500, the initial learning rate is set to 0.15, the update step is 300, the discount rate is 0.9, the attenuation rate is 0.001, the mini-batch capacity is 125, and the experience pool capacity is 3000.

Fig. 3 presents the detailed data concerning the time segments and price of time-of-use electricity price implemented in the region.

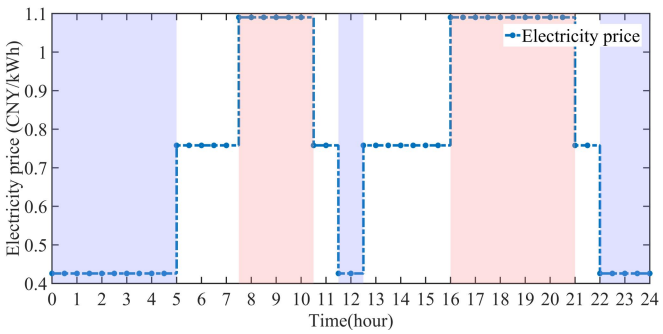

Fig.3. Time-of-use price of photovoltaic energy storage charging stations.

B. Optimization results and analysis

Fig. 4 depicts the reward obtained during the training process of the proposed enhanced DQN algorithm. During the early phase of the training, given the insufficiency of training samples, the agent actively explores the environment with a high learning rate. With the gradual accumulation of

samples, the reward curve climbs significantly and tends to converge. With the continuous increase of training rounds, the reward curve tends to be stable, and the agent successfully completes the learning of the optimal mapping relationship. Ensure that the decision of each agent remains reliable and stable in a dynamic, uncertain environment.

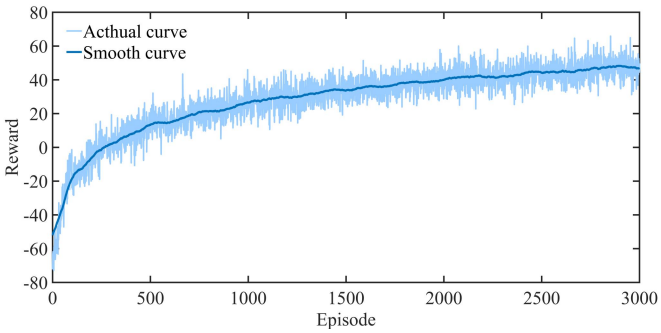

Fig. 4. Improved DQN algorithm training reward curve.

To validate the feasibility of the proposed approach and conduct an economic analysis of the photovoltaic and energy storage configuration for PESCS, four comparative schemes are established:

Scenario 1. Neither photovoltaics nor an energy storage system is configured.

Scenario 2. Configure photovoltaics; do not configure an energy storage system.

Scenario 3. Do not configure photovoltaics; configure an energy storage system.

Scenario 4 (This paper's scheme). Configure photovoltaics and energy storage systems at the same time.

The configuration outcomes are presented in TABLE 2. Across the four scenarios, the PESCS demonstrates favorable economic advantages compared to the conventional charging station. By comparing scenario 2 with scenario 3, it becomes apparent that the configuration benefit of PV is greater than that of the ESS configuration. However, solely configuring PV fails to realize peak-shaving and valley-filling.Therefore, the configuration of the ESS plays a crucial role in the PESCS. Thus, it is possible to configure PV as much as possible within a limited range and configure appropriate ESS to ensure a reliable and economical power supply. The economy of the PESCS has improved by 12.66%.

TABLE 2
CONFIGURATION RESULTS UNDER DIFFERENT SCHEMES.

Configuration	Scenario 1	Scenario 2	Scenario 3	Scenario 4
PV configuration capacity/ kW	–	396	–	396
Energy storage capacity/ kWh	–	–	600	628
Energy storage rated power/ kW	–	–	198	200
Average annual cost/ CNY	3239.2	2965.8	3186.8	2828.9

To attain the dual-win objective of maximizing the consumption of PV energy at the charging station and minimizing the charging expenses for the vehicle owners, a rational midday low service charge policy is established for the PESCS. This policy aims to steer EVs to modify their charging schedules. With the prerequisite of meeting the charging need, the peak charging period in the evening is transferred to the high photovoltaic output period at noon. The adjusted EV load of the EV is roughly bimodal, which reduces the pressure of no PV output in the evening peak.

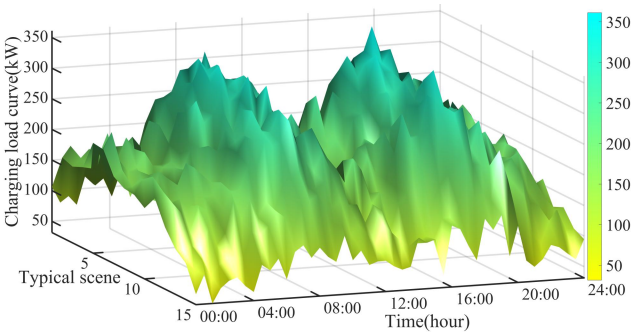

Fig. 5. Charging station load space-time distribution map.

Fig. 5 shows the variation trend of the load curve of the PESCS in typical scenarios. By understanding the change trend of the load curve of typical PESCS, a more economical energy storage configuration is formulated with a reasonable energy management strategy and considering the battery health status of ESS. After the PV and ESS are configured in the PESCS, their operation mode on typical days is further discussed. Its operation mode and energy exchange with the power grid are shown in Fig. 6, which shows the power balance state at each moment and the power supply source.

The RL algorithm is employed to address the model. When the installed capacity of the PV system installed in the charging station reaches 396 kW and the ESS power is 200 kW, the optimal capacity configuration of the ESS is determined to be 628 kWh. The corresponding PESCS has the largest net income. At this time, the proportion of PV supply charging load is 61.39%.

Taking into account the time-of-use price and load demand, during the period from 0:00 to 4:00 when the grid load reaches its trough and the electricity price is in the low-price range, the ESS gives priority to charging the EV load. Moreover, it charges the ESS based on the day's weather conditions. This approach notably cuts down the cost of electricity procurement.

From 8:00 to 11:00, when the charging demand of EVs reaches its peak, the charging price also enters the peak-rate period. During this time frame, the power generated by the PV system fails to fulfill the load requirements of EVs. The ESS discharges to make up for the power gap during the peak period and effectively avoids the cost pressure caused by the peak price.

During the period from 12:00 to 15:00, the surplus power generated by PV power generation is preferentially stored in the ESS. This ensures that the PESCS can obtain a reliable power supply through ESS when there is no PV power generation at night and the electricity price is at peak hours.

During the period from 18:00 to 21:00, the electricity price reaches its peak, and there is no output from the PV system. At this juncture, the load of EVs is primarily supplied by the ESS. When the ESS stored power is inadequate, the power grid steps in to provide supplementary power. For the rest of the periods with different electricity prices, the load demand is mainly met by the power grid, and the state of the ESS determines whether to participate in the discharge.

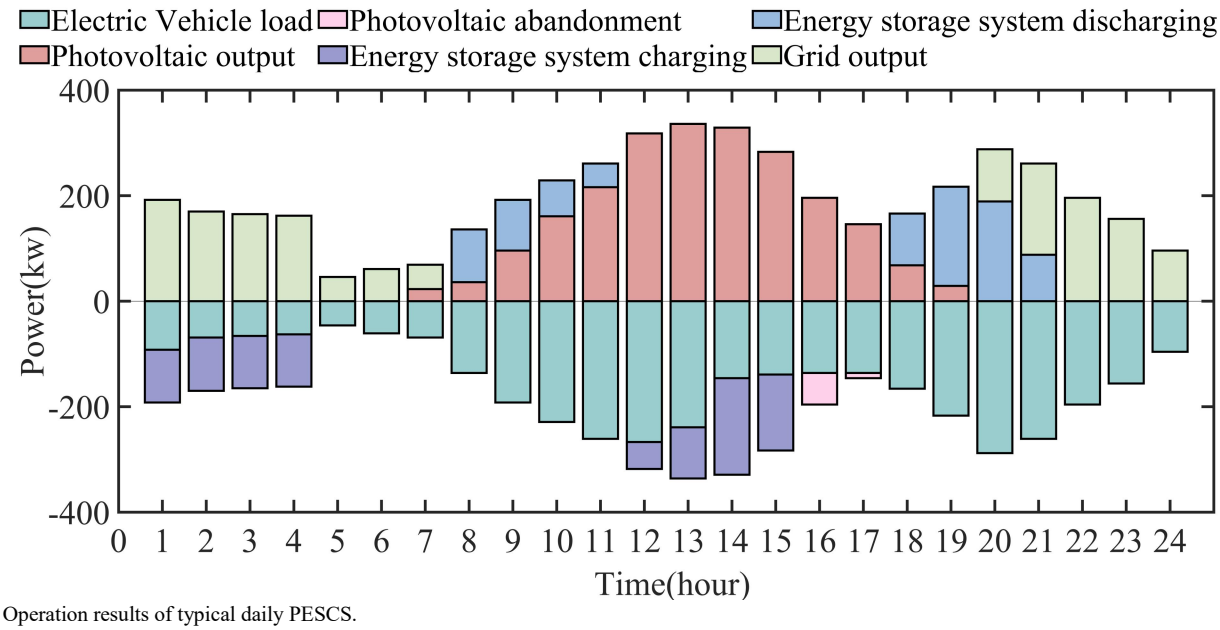

Fig.6. Operation results of typical daily PESCS.

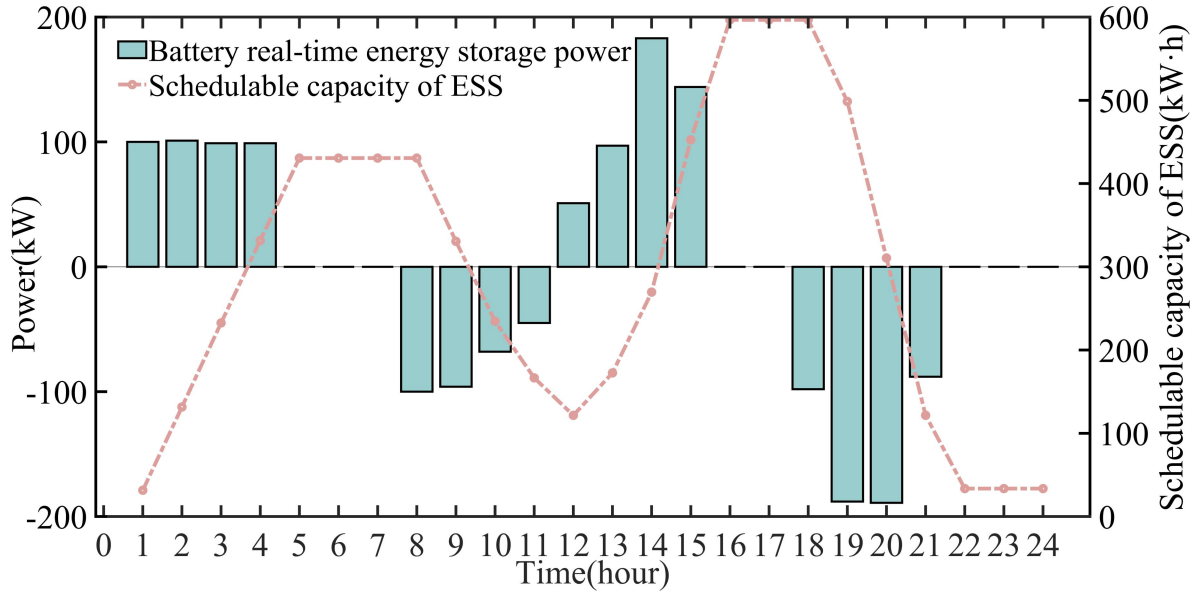

Fig.7. Charging and discharging power curve of ESS.

The detailed charging and discharging procedure of the ESS is presented in Fig. 7. In summary, the capacity configuration of the PESCS based on the Deep Q-Network algorithm can conform to the operating characteristics of each PESCS and collaboratively meet the load demand of EV in the station. Reasonable configuration of the ESS can not only promote the coordinated complementarity among PV, the ESS, and the power grid, but also effectively suppress the fluctuations of the load.

Through the analysis of the operation results, it is evident that the overall performance of the PESCS is superior under this configuration. Compared with before optimization, the ESS power curve is better smoothed. Under this configuration, the PV curtailment rate of PESCS is reduced to less than 3%, which proves that the configuration of ESS is reasonable and can effectively reduce PV curtailment.

For a long time, the state-of-health of the ESS battery has remained a bottleneck that restricts the rapid development of the PESCS. Therefore, the influence of state-of-health of health of the ESS on the economy of the PESCS is analyzed. Fig. 8 shows the influence of the ESS of the PESCS on the net income and comprehensive cost of PESCS with the change of service life.

It can be seen from Fig. 8 that the health of the ESS in the PESCS has a great influence on its net income and comprehensive cost. The healthier the ESS is, the longer the service life is, the lower the comprehensive cost of the PESCS is, and the higher the net income is. Therefore, formulating a reasonable configure the capacity of the PESCS is conducive to delaying the battery's state of health. This approach not only lowers the cost of ESS replacement but also improves the economic viability of the PESCS.

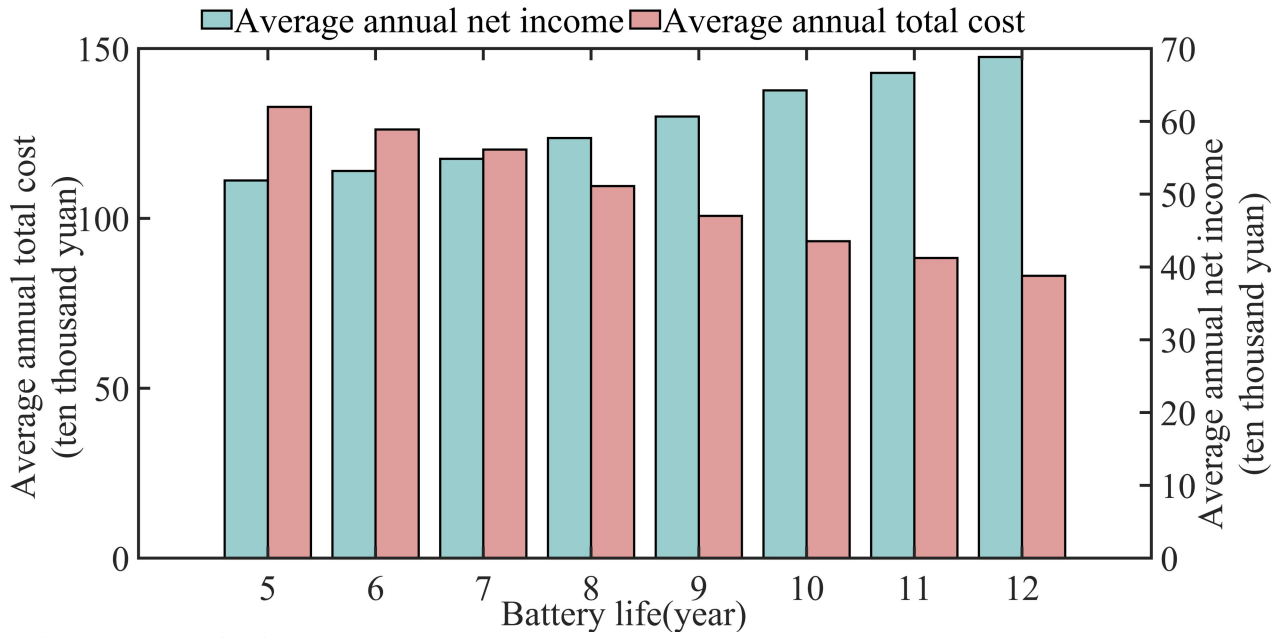

Fig. 8. The impact of battery health on PESCS.

Therefore, whether to consider the modeling of battery health status in the design stage of PESCS will significantly affect the net income of the PESCS. The traditional model adopts an empirical setting, which often leads to redundancy or deficiency of ESS capacity, and most traditional models have limited contribution to improving the net income of PESCS. Evidently, this paper takes into account the factors related to the battery's state-of-health during the design phase of the ESS. This approach offers notable benefits in ascertaining the optimal capacity configuration and enhancing the profitability of the PESCS in different years.

VI. Conclusion

The capacity configuration of a PESCS largely determines its operational mode and economic benefits. This paper puts forward a capacity configuration approach for PESCS relying on reinforcement learning algorithms. The approach takes into account the battery's state-of-health and makes use of the flexible complementary capacities of the ESS to improve the operational economic efficiency of the PESCS. Via case analysis, the following conclusions are reached:

(I) The DQN algorithm, relying on a time-of-use pricing energy management strategy, relieves the load demand of the PESCS and significantly improves its adjustable capacity during power shortages at peak electricity price periods through the ESS.

(II) The capacity configuration model of the PESCS comprehensively considers the battery's state-of-health, reduces ESS replacement costs, and increases the long-term economic benefits of the PESCS.

(III) The capacity configuration method for PESCS based on reinforcement learning, which combines the flexible complementary capability of ESS with the battery health state, can effectively enhance the economic efficiency of PESCS by up to 12.66%, with the PV abandoned rate being less than 3%.

The method put forward in this paper considers the battery health state, achieves the maximization of economic benefits for PESCS, and gives a reference for the improvement of PESCS capacity configuration and investment decision-making. However, the geographical layout of the ESS significantly impacts the overall economic benefits. Research in the future will aim at exploring how to select the actual locations of charging stations and the direction in which algorithms can be optimized.

References

[1] H. Lin, S. Liu, et al. "A two-stage robust optimal capacity configuration method for charging station integrated with photovoltaic and energy storage system considering vehicle-to-grid and uncertainty," Energy, Vol. 319, 135057, 2025.

[2] S. Wang, S. Bi, et al. "Reinforcement Learning for Real-Time Pricing and Scheduling Control in EV Charging Stations," IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp 849-859, 2021.

[3] Y. Luo, H. Hao, et al. "Multi-Objective Optimization of Integrated Energy Systems Considering Ladder-Type Carbon Emission Trading and Refined Load Demand Response," Journal of Modern Power Systems and Clean Energy, vol. 12, no. 3, pp 828-839, 2024.

[4] M. Massaoudi, H, Abu-Rub, et al. "Deep Learning in Smart Grid Technology: A Review of Recent Advancements and Future Prospects," IEEE Access, vol. 9, pp 54558-54578, 2021.

[5] S. Jang, A. Yoon, et al. "Optimal capacity determination of photovoltaic and energy storage systems for electric vehicle charging stations," Journal of Energy Storage, vol. 106, 114730, 2025.

[6] Y. Liu, P. Li, et al. "Research on Microgrid Superconductivity-Battery Energy Storage Control Strategy Based on Adaptive Dynamic Programming," IEEE Transactions on Applied Superconductivity, vol. 34, no. 8, pp 1-4, 2024.

[7] J. Zhang, L. Hou, et al. "Optimal operation of energy storage system in photovoltaic-storage charging station based on intelligent reinforcement learning. Energy and Buildings," Vol. 299, 113570, 2023.

[8] H. S. V. S. K. Nunna, A. Sesetti, et al. "Multiagent-Based Energy Trading Platform for Energy Storage Systems in Distribution Systems with Interconnected Microgrids," IEEE Transactions on Industry Applications, vol. 56, no. 3, pp 3207-3217, 2020.

[9] M. Shin, D. Choi, et al. "Cooperative Management for PV/ESS-Enabled Electric Vehicle Charging Stations: A Multiagent Deep Reinforcement Learning Approach," IEEE Transactions on Industrial Informatics, vol. 16, no. 5, pp 3493-3503, 2020.

[10] X. Dong, J. Shen, et al. "Simultaneous capacity configuration and scheduling optimization of an integrated electrical vehicle charging station with photovoltaic and battery energy storage system," Energy, vol. 289, 129991, 2024.

[11] X. Zheng, Y. Yao, "Multi-objective capacity allocation optimization method of photovoltaic EV charging station considering V2G," Journal of Central S.outh University, vol. 28, no. 2, pp 481-493, 2021.

[12] X. Dong, J. Shen, et al. “Simultaneous capacity configuration and scheduling optimization of an integrated electrical vehicle charging station with photovoltaic and battery energy storage system,” Energy, Vol. 289, 129991, 2024.
[13] Y. Liu, X. Han, et al. “Research on Control Strategy of Hybrid Superconducting Energy Storage Based on Reinforcement Learning Algorithm,” IEEE Transactions on Applied Superconductivity,vol. 34, no. 8, pp 1-4, 2024.
[14] J. Li, S. Yu, “Learning Path Recommendation Based on Reinforcement Learning,” Engineering Letters, vol. 32, no. 9, pp 1823-1832, 2024.
[15] F. Härtel, T. Bocklisch, “Minimizing Energy Cost in PV Battery Storage Systems Using Reinforcement Learning,” IEEE Access, vol. 11, pp 39855-39865, 2023.
[16] Y. Liu, Z. Jiang, et al. “Supply and demand balance control strategy for microgrids considering uncertainty and disturbance based on online H∞ policy iteration,” Journal of Computing and Information Science in Engineering, vol. 24,no. 9, 090902, 2024.
[17] T. Sun, C. Ma, “Cloud Computing-based Parallel Deep Reinforcement Learning Energy Management Strategy for Connected PHEVs,” Engineering Letters, vol. 32, no. 6, pp 1210-1220, 2024.
[18] D. Yu, R. Tang, “Optimal allocation of photovoltaic energy storage in DC distribution network based on interval linear programming,” Journal of Energy Storage, Vol. 85, 110981, 2024.
[19] F. Jiang, K Xue, et al. “Bi-level Optimal Configuration of Energy Storage System in Distribution Network Using an Improved Grey Wolf Optimization Algorithm,” IAENG International Journal of Applied Mathematics, vol. 55, no. 3, pp 582-593, 2025.
[20] G. Chen, J. Li, et al. “Optimal Configuration of Renewable Energy DGs Based on Improved Northern Goshawk Optimization Algorithm Considering Load and Generation Uncertainties,” Engineering Letters, vol. 31, no. 2, pp 511-530, 2023.
[21] Y. Wang, S. Guo, et al. “A comprehensive review of machine learning-based state of health estimation for lithium-ion batteries: data, features, algorithms, and future challenges,” Renewable and Sustainable Energy Reviews, vol. 224, 116125, 2025.

YANG LIU received the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2018. He is currently an Associate Professor with the School of Electrical Engineering, Shenyang University of Technology, Shenyang. His current research interests include adaptive dynamic programming, reinforcement learning, and integrated energy system operation optimization.

MINGDA SONG was born in Dalian City, Liaoning Province, China, in 1999. He obtained a Bachelor's degree in Electrical Engineering from Liaoning Technical University in 2022. He is currently pursuing the M.S. degree in Electrical Engineering at Shenyang University of Technology, Liaoning, P. R. China. His main research areas include the application of reinforcement learning in integrated energy system operation optimization.

HENGYU LIU received the M.S. degree in electrical engineering automation from Northeastern University, China, in 2018, where he is currently pursuing the Ph.D. degree with the School of Shenyang University of Technology. Since 2018, he has been with Electric Power Research Institute of State Grid Liaoning Electric Power Co., Ltd. His research interests include comprehensive energy technology and virtual power plant technology.

ZUOXIA XING received the Ph.D. degree in automation of electric power systems from Beijing Jiaotong University, Beijing, China, in 2008. She has been a Professor with the Shenyang University of Technology since 2018. Her current research interests include new energy control, multi-energy complementarity, energy storage, and consumption.

Qiwen Li was born in Changde City, Hunan Province, China, in 2001. He obtained a Bachelor's degree in Electrical Engineering from Shenyang University of Technology in 2023. He is currently pursuing the M.S. degree in Electrical Engineering at Shenyang University of Technology, Liaoning, P. R. China. Her research direction is new energy control.

An Zhu was born in Fuxin City, Liaoning Province, China, in 2000. She obtained a Bachelor's degree in Electrical Engineering from Liaoning Technical University in 2022. She is currently pursuing the M.S. degree in Electrical Engineering at Shenyang University of Technology, Liaoning, P. R. China. Her research direction is integrated energy.