Article:
Xing, C, Li, K orcid.org/0000-0001-6657-0522 and Su, J (2023) Thermal Constrained Energy Optimization of Railway Co-phase Systems with ESS Integration - An FRA-pruned DQN Approach. IEEE Transactions on Transportation Electrification, 9 (4). 5122 -5139. ISSN 2332-7782

https://doi.org/10.1109/TTE.2022.3218762

Thermal Constrained Energy Optimization of Railway Co-phase Systems with ESS Integration - An FRA-pruned DQN Approach

Chen Xing, Kang Li, *Senior member, IEEE,* and Jialei Su

***Abstract*—This paper investigates the railway co-phase traction power supply system (TPSS) with a power flow controller (PFC) to address the power quality and neutral section issues. To collect the regenerative energy and achieve a more flexible power flow, the energy storage system (ESS) is integrated into the co-phase system. As the key components, the reliability of power electronics modules in PFC and battery cells in ESS is highly related to their thermal performance. It is therefore vital to consider their operational thermal dynamics, leading to the proposal of a deep Q network (DQN) based thermal constrained energy management strategy in this paper. Firstly, the system power flow model and electrothermal models for power electronics modules and battery cells are all established. Then, a DQN method is adopted to learn an optimized policy for peak power shaving while meetings thermal constraints. Finally, an FRA-based pruning method is proposed to reshape the agent to become more compact without sacrificing its performance. Case studies confirm that the proposed strategy can effectively reduce the peak traction power supply by up to 42.0%, and achieve up to 94.1% thermal reduction. The FRA-based pruning can achieve up to 89.9% agent size reduction and 87.2% computation savings.**

Index Terms - Co-phase railway traction system, thermal constraint, deep Q-learning network, agent pruning.

NOMENCLATURE

A. Abbreviations (Alphabetically)

DNN	Deep neural network
DQN	Deep Q-learning network
DRL	Deep reinforcement learning
EMU	Electrical multi units
ESS	Energy storage system
FRA	Fast recursive algorithm
HESS	Hybrid energy storage system
IGBT	Insulated gate bipolar transistor
MDP	Markov decision process
MILP	Mixed integer linear programming
MMC	Modular multilevel converter
MPC	Model predictive control
PFC	Power flow controller
SoC	State of charge
TPSS	Traction power supply system

This work was supported in part by the Energy Hub Project Funds from Ofgem. (Corresponding author: Kang Li.)

Chen Xing, Kang Li, and Jialei Su are with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, LS2 9JT, UK.

Chen Xing: elcx@leeds.ac.uk

Kang Li: K.Li1@leeds.ac.uk

Jialei Su: eljs@leeds.ac.uk

B. Parameters and Variables (Alphabetically)

C_1, C_2	Battery internal and shell thermal capacity
$C_{E.T}, C_{E.D}$	Temperature coefficients of switching loss
C_{ESS}	Rated capacity of ESS
f_{sw}	Switching frequency
$I_{ref}, V_{ref}, T_{ref}$	Reference current, voltage, and temperature
K_v, K_i	Correction coefficient of voltage and current for power loss
k_1, k_2	Heat conduction coefficients
P_a, Q_a	Active and reactive power from phase a
$P_{ca}, Q_{ca}, P_{cb}, Q_{cb}$	Active and reactive compensation power of converters A and B
P_{con}, P_{sw}	Conduction loss and switching loss
P_{ESS}	Charging/discharging p ower of ESS
S_L, P_L, Q_L, φ_L	Apparent power, active power, reactive power and phase angle of traction loads
T_{amb}	Ambient temperature
T_{in}, T_{sh}	Battery internal and shell temperature
T_j, T_h	Junction temperature and heat sink temperature

I. INTRODUCTION

THE urbanization and industrialization trends worldwide have significantly increased the public transport demands, and the railway systems are playing an even more important role in urban and city transportation. Meanwhile, railway electrification has attracted much attention in recent years as a key part of the overall transportation decarbonization effort. Given the landscape of a significant increase in traction energy demand and the challenges of the overall decarbonization targets, it is vital to guarantee the traction power quality and enhance the stability of the power supply system [1].

The traditional single-phase 50Hz/27.5kV AC traction power supply system is widely deployed [2]. However this traditional power supply system may raise some concerns such as the three-phase imbalance, reactive power and neutral section issues [3]. To tackle these issues, a novel co-phase railway power supply system as shown in Fig. 1 has been proposed and widely studied [4], [5], where a back-to-back converter acts as a PFC to compensate reactive power and balance the three-phase power grid. Meanwhile, the neutral sections can be eliminated. In [6], a 10MVA-rated co-phase system has been operated in China. Furthermore, braking of electrical multi-units (EMUs) will produce a large amount of regenerative energy which is undesirable to feed them back to the three-phase power grid. On the other hand, ESS is widely

applied to collect regenerative energy and further provide the ancillary services such as spinning reserve, load shifting and power compensation [7]–[9]. For example, in [7], the ESS is integrated into the railway co-phase systems to obtain a minimum daily comprehensive cost and maximum regenerative utilization. In [8], [9], the hybrid energy storage system (HESS) is also integrated into the railway power supply system to smooth the fluctuations of traction/regenerative power. In this paper, the ESS is integrated to regulate the high-power, dramatically-varying traction loads so as to achieve peak traction power shaving and smooth acceptance of regenerative power with the aid of the smart energy management strategy. The co-phase system with ESS integration is shown in Fig. 1.

As more power electronics modules and ESS have been applied in traction power supply systems for improving power quality, integrating renewables, smoothing traction load, and absorbing regenerative power, their lifespans may be significantly reduced due to poor thermal management, even leading to more frequent system failure events [10], [11]. For example, as reported in [10], the lifespan of the ESS will be reduced by two months per degree rise when the temperature is higher than 40°C. As demonstrated in [11], the failure rates of power converters are higher than those of other devices in railway power supply system, and more than 50% of the converter failures are caused by the insulated gate bipolar transistors (IGBTs) and freewheeling diodes. Thus it is vital to enhance the reliability and extend the lifespan of the ESS and power electronics modules. The ESS and PFC are among the key components of the railway co-phase systems as illustrated in Fig. 1 to serve for the high voltage and high power traction loads, and the co-phase power supply system operation performance highly depends on the reliability of battery cells and power electronics modules including the IGBTs and diodes. Some researchers have discussed the reliability enhancement of ESS integrated into the railway system. For example, in [7], ESS degradation is considered to quantify the operation cost of the railway co-phase system. By optimizing the operation cost, the lifespan of ESS can be prolonged. In [3], a mixed-integer linear programming (MILP) optimization model with ESS degradation considered is formulated for the operation of the railway ESS-integrated co-phase system. However, as illustrated in [12], [13], besides the dynamics of the electrical systems, the failure rates of the battery cells and power electronics modules are highly related to the internal temperatures of the battery cells and the junction temperatures of the power electronics modules respectively. To control the thermal dynamics of the battery cells and power electronics modules, besides taking thermal capacity into consideration in the system design or applying an advanced cooling system [14]–[16], to develop a suitable thermal management strategy is an effective approach with the advantage that extra investment may not be required and several thermal management strategies for the ESS and power electronics modules have been proposed. For the ESS, in [17], a multi-objective optimization framework is developed to optimize the charging patterns of the ESS in EVs with considering the temperature variations of the battery cells. In [10], [18], two energy management strategies using deep reinforcement learning (DRL) with battery thermal constraint for the hybrid electric buses are proposed. For the power electronics modules, a thermal balancing method by controlling the capacitor voltage of each submodule in the modular multilevel converter (MMC) is proposed in [19]. Furthermore, in [12], the power electronics module lifetime affected by the junction temperature is considered in the co-phase railway PFC control. By optimizing the single-phase space vector pulse width modulation (SVPWM), the average temperature and temperature fluctuation of the power electronics modules are reduced. In [11], a coordinated control strategy to alleviate thermal stress of the power electronics modules in PFC is proposed and applied in the railway co-phase system. However, the aforementioned thermal management strategies for the power electronics modules mainly focus on individual component-level control rather than system-level energy management. Few researchers have considered thermal management based on the electrothermal models for both the ESS and power electronics modules in the railway power supply system. Given that these two key devices are strongly coupled during the operation, it is vital to develop a unified energy management strategy for the railway co-phase system to optimize the power flow while meeting thermal constraints for both ESS and PFC.

DRL is an effective machine learning approach for decision-making and optimization problems, and it has attracted much attention due to its model-free features [20]. A large amount of studies have focused on the application of the DRL in intelligent transportation [10], [21]–[23], and smart energy management [24], [25]. For example, in [10], a knowledge-based, multiphysics-constrained energy management strategy for hybrid electric buses is proposed, with an emphasized consciousness of both thermal safety and degradation of the onboard lithium-ion battery (LIB) system. In [21], [22], two DRL methods have been developed respectively for optimizing the fast charging procedure of the electric vehicles. In [23], a battery health-aware and DRL-based energy management framework is proposed for power-split hybrid electric vehicles in a naturalistic driving scenario. In [25], a model-free DRL method is proposed to optimize the battery energy arbitrage considering an accurate battery degradation model. Compared to the model-based methods, the DRL methods have demonstrated several distinctive advantages: 1) DRL algorithms are model-free and self-adaptable, and can learn the knowledge from historical data; 2) DRL algorithms can learn a good policy even under a very complex nonlinear environment with the aid of deep neural networks (DNNs). In this study, the thermal management for both the ESS and power electronics modules are considered and the formulated problem for optimizing the energy flow of the railway co-phase system is a complex nonlinear optimization problem subject to a number of nonlinear constraints. The linear programming techniques are difficult to solve this problem unless the nonlinear terms in the model are linearized. However linearization will inevitably reduce the model accuracy. The distinctive features of the DRL approaches offer great potentials to learn a smart energy management strategy from the complex nonlinear operational environment of the ESS integrated co-phase system. Even so, the aforementioned DRL-based methods still suffer from a

bottleneck in real-life applications. That is the agent realized by the DNN requires large computation and memory resources, which limits its deployment in systems that have limited hardware resources [26]. Thus the agents trained by DNN-based DRL need to be more compact without sacrificing their performance.

This paper aims to bridge the aforementioned research gaps and propose a novel unified FRA-pruned DQN-based thermal constrained energy management strategy to shave the peak traction power with the aid of the ESS, and reduce the reactive power injection into the three-phase power grid as much as possible. The main contributions are summarized as follows:

1) The model of the co-phase system with ESS integration is established. The power flow of the co-phase system is analysed and the compensation power of the PFC is derived based on the power of the ESS and phase angle of the three-phase power grid.
2) The electrothermal models of the ESS and power electronics modules in the PFC are built respectively. The thermal dynamics can be analysed using the electrothermal models based on the compensation power of PFC and the power of ESS.
3) With the electrothermal models, an intelligent unified DQN-based energy management agent is developed to optimize the power flow of the co-phase system, so as to achieve the peak traction load shaving while meeting the thermal constraints for both ESS and power electronics modules.
4) A novel FRA-based pruning method for the DNN-based DQN agent is proposed. After reshaping, the agent becomes more compact with the redundant neurons in the DNN being removed. While the performance is guaranteed, and the computational complexity and required memory resources of the agent are significantly reduced.

The remainder of the article is organized as follows. In Section II, the power flow model of the ESS integrated co-phase system is developed. The electrothermal models of ESS and power electronics modules in the PFC are also detailed. Section III formulates the optimization problem of thermal constrained energy management. The DQN-based thermal constrained energy management strategy is developed in Section IV. The FRA-based pruning method for the DQN agent is proposed in Section IV. Section V presents the case studies to verify the effectiveness of the proposed strategy. Section VI concludes this paper.

II. Model Formulation

A. Power Flow Analysis of ESS Integrated Co-phase System

As illustrated in Fig. 1(a), the railway power supply from the three-phase the grid is split into two single-phase channels by the V/V transformer, which is a traditional transformer configuration widely adopted in the railway traction system. One single-phase a connects to the traction line directly to feed traction loads. The back-to-back PFC is connected between phase bc and phase ac, which can balance the three-phase power grid, compensate the reactive power and suppress the harmonics. The configuration of PFC with ESS integration is shown in Fig. 1(b). The MMC is adopted for the high power compensation. The ESS is integrated into the co-phase system by connecting a set of battery cells with the submodules of MMC through a DC-DC converter. The ESS can achieve more flexibility for power flow management and provide ancillary services. In this paper, in the premise of ensuring the balance of the three-phase power grid, the ESS is mainly utilized to achieve the load shifting so as to shave the traction/ regenerative peak power and smooth the traction loads.

The traction loads are inductive, which consume both active power and reactive power. Based on Fig. 1(a), one part of the traction loads is directly fed by the phase a of the power grid, and the other part is compensated by the PFC. Thus (1) can be obtained.

$$\begin{cases} P_L = S_L \cdot \cos\varphi_L = P_{ca} + P_a \\ Q_L = S_L \cdot \sin\varphi_L = Q_{ca} - Q_a \end{cases} \tag{1}$$

where S_L, P_L, Q_L and φ_L represent the apparent power, active power, reactive power and phase angle of the traction loads respectively. P_{ca} and Q_{ca} represent the active and reactive compensation power from converter A of PFC. P_a and Q_a represent the active and reactive power fed by the phase a.

With the integration of the ESS, the traction active power can be compensated by both the three-phase power grid and the ESS. The phasor diagram of the power in the co-phase system can be shown in Fig. 2. Based on the power balance principle and phasor diagram in Fig. 2, the following equation is established

$$\begin{cases} P_a = P_L - P_{ca} \\ Q_a = \tan\left(\frac{\pi}{6} - \varphi\right) P_a \\ P_{ca} = P_{cb} + P_{ESS} \\ Q_{ca} = Q_L + Q_a \\ P_{cb} = P_{cb} \\ Q_{cb} = \tan\left(\frac{\pi}{6} + \varphi\right) P_{cb} \end{cases} \tag{2}$$

where P_{ESS} represents the power of ESS. P_{cb} and Q_{cb} represent the active and reactive compensation power of converter B in PFC. φ denotes the three-phase angle. Here, the three-phase angles φ_a, φ_b and φ_c are all set to φ to guarantee the full balance of the three-phase power grid.

Furthermore, since the vector sum of the apparent power of the three-phase power grid equals zero, i,e. $\tilde{S}_a + \tilde{S}_b + \tilde{S}_c = 0$, the relationship between P_{cb} and P_L, P_{ESS} can be derived

Then the compensation power of PFC can be expressed as (4) by substituting (3) into (2). It means that, once the power of ESS and phase angle of the power grid are determined, the corresponding compensation power of PFC is determined accordingly based on the traction loads.

$$\begin{cases} P_{ca} = \alpha P_L + (1-\alpha) P_{bat} \\ Q_{ca} = Q_L + \tan\left(\frac{\pi}{6} - \varphi\right)(1-\alpha)(P_L - P_{bat}) \\ P_{cb} = \alpha (P_L - P_{bat}) \\ Q_{cb} = \tan\left(\frac{\pi}{6} + \varphi\right) \alpha (P_L - P_{bat}) \end{cases} \tag{4}$$

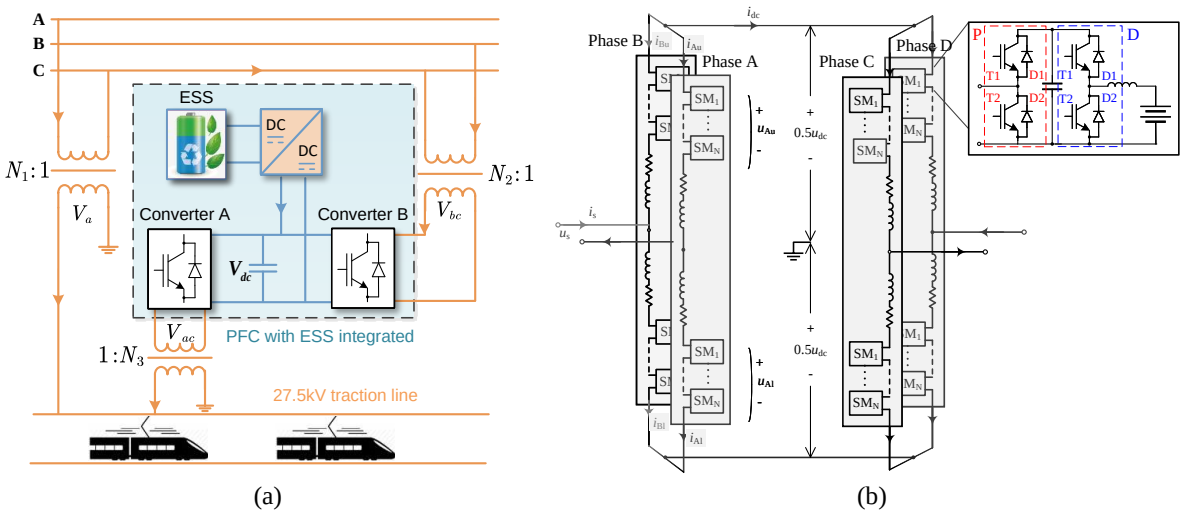

Fig. 1. Railway co-phase system with ESS integration. (a) System structure; (b) Configuration of MMC-based PFC with ESS integration.

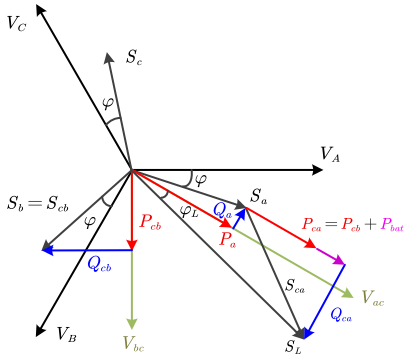

Fig. 2. Phasor diagram of ESS integrated co-phase system.

B. Electrothermal Model of Power Electronics modules in PFC

1) Modeling of PFC and DC-DC converter: The back-to-back MMC-based PFC with ESS integration as shown in Fig. 1(b) can be seen as the combination of two single-phase MMCs labeled as converters A and B respectively. Both single-phase MMCs act as the inverter and rectifier respectively under the specific operation modes. As shown in Fig. 1(b), each submodule in the single-phase MMC includes four IGBTs and four diodes. Two sets of IGBTs and diodes in Fig. 1(b), labeled as P-T1, P-D1, P-T2, and P-D2, mainly perform the function of PFC such as power compensation. Thus the converter made of P-T1, P-D1, P-T1, and P-D2 is called as PFC converter for convenience. Another two sets of IGBTs and diodes, labeled as D-T1, D-D1, D-T2, and D-D2, constitute the DC-DC converter to connect the battery cells with the PFC.

In the following, the modeling of PFC and DC-DC converters is discussed separately. To facilitate the analysis, take the inverter mode of the single-phase MMC as an example. When the single-phase MMC is operated in the inverter mode, the AC side current is evenly distributed through the upper and lower arms. Assume the AC side voltage u_s and current i_s as:

$$\begin{cases} u_s = U_m \sin(\omega t) \\ i_s = I_m \sin(\omega t - \phi) \end{cases} \tag{5}$$

where U_m and I_m are the maximum amplitudes of the voltage and current of the AC side. ω is the AC frequency and ϕ is the phase angle of the AC side in the MMC. Based on the aforementioned analysis, when the compensation power of the PFC is determined, I_m can be easily obtained based on the compensation power and port voltages of the PFC.

As shown in Fig. 1(b), taking phase A of the single-phase MMC as an example, the voltages u_{Au}, u_{Al} and currents i_{Au}, i_{Al} of the upper arm and lower arm of the phase A can be expressed as

$$\begin{cases} u_{Au} = \frac{U_{dc}}{2} - u_s, \ u_{Al} = \frac{U_{dc}}{2} + u_s \\ i_{Au} = \frac{I_{dc}+i_s}{2}, \ i_{Al} = \frac{I_{dc}-i_s}{2} \end{cases} \tag{6}$$

where U_{dc} and I_{dc} represent the voltage and current of the DC side.

Since the active power on the DC side should be equal to that from the AC side, it yields

$$\begin{aligned} U_{dc} I_{dc} &= \frac{U_m}{\sqrt{2}} \cdot \frac{I_m}{\sqrt{2}} \cdot \cos\phi \\ &= \frac{1}{2} U_m I_m \cos\phi = \frac{1}{4} m U_{dc} I_m \cos\phi \end{aligned} \tag{7}$$

where $m = \frac{2U_m}{U_{dc}}$ is the modulation ratio. Thus it has

$$I_{dc} = \frac{1}{4} m I_m \cos\phi \tag{8}$$

Substitute (5) and (8) into (6), then

$$\begin{cases} i_{Au} = \frac{1}{8} m I_m \cos\phi + \frac{1}{2} I_m \sin(\omega t - \phi) \\ i_{Al} = \frac{1}{8} m I_m \cos\phi - \frac{1}{2} I_m \sin(\omega t - \phi) \end{cases} \tag{9}$$

According to [27], the duty ratio of the upper and lower arms can be given as

$$\begin{cases} n_{Au} = \frac{1}{2}(1 - m\sin(\omega t)) \\ n_{Al} = \frac{1}{2}(1 + m\sin(\omega t)) \end{cases} \tag{10}$$

$$P_{cb}=\alpha(P_L-P_{ESS})=\frac{\cos\left(\frac{\pi}{6}+\varphi\right)\left(\sin\varphi\cdot\sin\left(\frac{\pi}{6}+\varphi\right)+\cos\varphi\cdot\cos\left(\frac{\pi}{6}+\varphi\right)\right)(P_L-P_{ESS})}{\sin\varphi\left(\cos\left(\frac{\pi}{6}+\varphi\right)\sin\left(\frac{\pi}{6}+\varphi\right)-\cos\left(\frac{\pi}{3}+\varphi\right)\sin\left(\frac{\pi}{3}+\varphi\right)\right)+\cos\varphi\left(\cos\left(\frac{\pi}{6}+\varphi\right)^2+\sin\left(\frac{\pi}{3}+\varphi\right)^2\right)} \tag{3}$$

To further estimate the thermal stresses of each power electronics module in PFC, the average and root-mean-square (RMS) currents of the IGBTs and diodes need to calculate. Take the upper arm as an example, assume the AC side current reaches zero when the phase angles are θ, $\pi+2\phi-\theta$, $2\pi+\theta$ within a fundamental-frequency period. The working time is $(\pi+2\phi-\theta, 2\pi+\theta)$ for P-T1 and P-D2, and $(\theta, \pi+2\phi-\theta)$ for P-T2 and P-D1. Therefore, the average and RMS currents of P-T1 can be expressed as

$$i_{P-T1.ave}=\frac{1}{2\pi}\int_{\pi+2\phi-\theta}^{2\pi+\theta} n_{Au}i_{Au}d(\omega t) \tag{11}$$

$$i^2_{P-T1.rms}=\frac{1}{2\pi}\int_{\pi+2\phi-\theta}^{2\pi+\theta} n_{Au}i^2_{Au}d(\omega t) \tag{12}$$

The average and RMS currents of P-T2 and P-D1, P-D2 can be also calculated based on (9) and (10). The calculation equations have been detailed in [27], and are therefore omitted in this paper.

Besides, for the devices D-T1, D-T2 and D-D1, D-D2 used for the DC-DC conversion, the average and RMS currents are mainly related to the duty ratio D, output current I_0 and current ripple ratio r. The corresponding relationships can be expressed as

$$i_{D-T1,2/D-D1,2.ave}=f^{DC}_{ave}(D,I_0) \tag{13}$$

$$i_{D-T1,2/D-D1,2.rms}=f^{DC}_{rms}(D,I_0,r) \tag{14}$$

where $D=1-\frac{V_I}{V_o}$ when the DC-DC converter is operated at the Boost mode and $D=\frac{V_o}{V_I}$ at the Buck mode. V_I and V_o are the port voltages of the DC-DC converter. The output current I_0 can be calculated based on the battery power P_{ESS} and the DC voltage V_o. f^{DC}_{ave} and f^{DC}_{rms} define the relationships between the average and RMS currents and the parameters D, I_0 and r, and please refer to [28] for details.

2) Loss model: The losses of the power electronics modules include conduction loss and switching loss. The conduction loss can be estimated as follows:

The conduction losses of IGBTs and diodes are determined by the voltage drop in IGBTs/diodes and the current flowing through them. The voltage drop V_{ce} of IGBT or diode at the junction temperature T_j can be expressed as:

$$V_{ce}(T_j)=V_{ce0}(T_j)+R_{ce0}(T_j)I_C \tag{15}$$

By fitting the relationship between $V_{ce}(T_j)$ and I_C at the temperatures T_u and T_l (generally $T_u=125°$C and $T_l=25°$C), two sets of parameters ($V_{ce0}(T_u)$, $R_{ce0}(T_u)$) and ($V_{ce0}(T_l)$, $R_{ce0}(T_l)$) can be obtained. Generally speaking, both sets of parameters are provided by the official datasheets of the power electronics modules. The function of $V_{ce0}(T_j)$ and $R_{ce0}(T_j)$ at T_j can be approximated by interpolation as follows:

$$V_{ce0}(T_j)=(V_{ce0}(T_u)-V_{ce0}(T_l))(T_j-T_l)/100+V_{ce0}(T_l) \tag{16}$$

$$R_{ce0}(T_j)=(R_{ce0}(T_u)-R_{ce0}(T_l))(T_j-T_l)/100+R_{ce0}(T_l) \tag{17}$$

Thus the conduction loss of IGBT or diode is given as:

$$\begin{aligned}P_{con.T/D}(T_j,I_{ave},I_{rms})&=\frac{1}{T_j}\int_0^T V_{ce}(T_j)I_C dt\\&=V_{ce0}(T_j)|I_{ave}|+R_{ce0}(T_j)I^2_{rms}\end{aligned} \tag{18}$$

where I_{ave} and I_{rms} can be obtained based on (11)-(12) for the IGBT and (13)-(14) for the diode.

Besides, at the switching instant, the operations of turning on/off for IGBTs and diodes will produce the switching losses which can be estimated by:

$$\begin{aligned}P_{sw.T}=&f_{sw}E_{on/off}\left(\frac{I_{ave}}{I_{ref}}\right)^{K_{i.T}}\left(\frac{V_{DC}}{V_{ref}}\right)^{K_{v.T}}\\&\cdot(1+C_{E.T}(T_j-T_{ref}))\end{aligned} \tag{19}$$

$$\begin{aligned}P_{sw.D}=&f_{sw}E_{rr}\left(\frac{I_{avg}}{I_{ref}}\right)^{K_{i.D}}\left(\frac{V_{DC}}{V_{ref}}\right)^{K_{v.D}}\\&\cdot(1+C_{E.D}(T_j-T_{ref}))\end{aligned} \tag{20}$$

where f_{sw} is switching frequency, I_{ref} , V_{ref} , and T_{ref} are the reference current, voltage, and temperature respectively, when the switching loss is measured as given in the datasheet. K_v and K_i are the correction coefficient for voltage and current dependency of the switching loss. $C_{E.T}$ and $C_{E.D}$ are the temperature coefficients of switching loss for IGBT and diode respectively.

In summary, the total average losses of the IGBT or diode are given as

$$P_{T.IGBT/D}=P_{con.T/D}+P_{sw,T/D} \tag{21}$$

3) Thermal estimation: The equivalent thermal networks of the power electronics module can be divided into two types either a Cauer model or a Foster model. Based on [27], for the Cauer model, the material property and geometry of the device are normally required to calculate the related parameters, which in many cases are difficult to obtain. While the parameters of the Foster model are easy to identify and usually provided in the datasheet. Therefore, the Foster model, as shown in Fig. 3, is used as the equivalent thermal network to estimate the junction temperature which is given by

$$T_{j.IGBT}=P_{T.IGBT}Z_{j-h.IGBT}+T_h \tag{22}$$

$$T_{j.D}=P_{T.D}Z_{j-h.D}+T_h \tag{23}$$

$$T_h=(P_{T.IGBT}+P_{T.D})Z_{h-a}+T_a \tag{24}$$

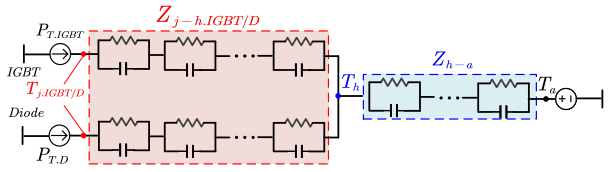

Fig. 3. Thermal network of power electronics modules.

where $T_{j.IGBT}$ and $T_{j.D}$ represent the junction temperatures of the IGBT and diode respectively. T_h represents the temperature of the heat sink. T_a represents the ambient temperature. $Z_{j-h.D}$ and $Z_{j-h.IGBT}$ represent the equivalent thermal networks of the IGBT and diode between the junction and heat sink. And Z_{h-a} represents the equivalent thermal network between the heat sink and ambient.

The thermal network in Fig. 3 reflecting the thermal dynamics can be expressed by

$$Z = \sum_{i=1}^{n} R_{th\_i}\left(1 - e^{-\frac{t}{\tau_{th\_i}}}\right) \tag{25}$$

where n is the order number of the thermal model and R_{th_i} is the thermal resistance. τ_{th_i} is the time constant of each RC network depending on R_{th_i} and C_{th_i}, and the relationship is given by

$$\tau_{th\_i} = R_{th\_i} \times C_{th\_i} \tag{26}$$

C. Electrothermal Model of ESS

The ESS is an important component in achieving more flexibility in the power flow. In this part, an electrothermal model of the ESS is adopted. The temporal evolution of the state of charge (SoC) can be formulated in a discrete form

$$SoC\left(k+1\right) = SoC\left(k\right) + \frac{I\left(k\right) \cdot T_s}{3600 C_{ESS}} \tag{27}$$

where $I\left(k\right)$ represents the charging/discharging current at time step k. C_{ESS} is the rated capacity of the ESS. T_s is the sampling time.

To model the electrothermal characteristics of the battery cells in ESS, a transient-state equivalent electric circuit model is adopted to represent the battery cell, which is shown in Fig. 4. This first-order simplified thermal model is widely adopted in the battery thermal modeling field [29], where the OCV is the battery open-circuit voltage, R represents the battery internal resistance. The RC network to capture the battery relaxation process is set to a constant network since the model accuracy can be guaranteed without being highly affected by factors such as SoC, temperature and current [29]. V and I are the terminal voltage and current of the battery cell. The reference directions of the current and voltage are shown in Fig. 4. When the actual directions of the current and voltage are in alignment with the labeled reference directions, the battery is operated in the discharging state and the output power is positive correspondingly. Conversely, when the actual direction of the current is opposite to its referenced direction, the battery is in the charging state and the output power is negative.

Based on [29], battery OCV is mainly dependent on the battery SoC and insensitive to the temperature variations, e.g. less than 10mV as temperature changes from $-10°$C to $50°$C. Also as stated in [30], the variation of battery OCV in terms of ambient temperature is negligible in the middle and high ranges of SoC (from 20% to 100% SoC) especially when the temperature is higher than $0°$C. Thus the battery OCV can be expressed as follows:

$$OCV\left(k\right) = f_{OCV}\left(SoC\left(k\right)\right) \tag{28}$$

The battery internal resistance R depends on the battery internal temperature as expressed in (29).

$$R\left(k\right) = f_R\left(T_{in}\left(k\right)\right) \tag{29}$$

Suppose the over-potential across the RC network shown in Fig. 4 is v_1, and that the load current keeps constant during the sampling period. Based on the dynamics of the RC network, (30) is given.

$$V_1\left(k+1\right) = a_1 V_1\left(k\right) + b_1 I\left(k\right) \tag{30}$$

where a_1 and b_1 are the electrical circuit model parameters identified by curve fitting the experimental data.

The terminal voltage of the battery cell can be given as

$$V\left(k\right) = f_{OCV}\left(SoC\left(k\right)\right) + f_R\left(T_{in}\left(k\right)\right) \cdot I\left(k\right) + V_1\left(k\right) \tag{31}$$

Based on the principle of power balance for the circuit shown in Fig. 4, the charging/discharging power of the battery cells can be given by

$$P_{ESS}\left(k\right) = I\left(k\right) OCV\left(k\right) - R\left(k\right) I^2\left(k\right) - I\left(k\right) V_1\left(k\right) \tag{32}$$

where $P_{ESS}\left(k\right)$ is the charging/discharging power at time step k. $P_{ESS}\left(k\right) < 0$ represents the charging to the ESS, $P_{ESS}\left(k\right) > 0$ represents the discharging from the ESS. Then the charging/discharging current can be obtained by solving (32), yielding

$$I = \frac{\left(OCV - V_1\right) - \sqrt{\left(OCV - V_1\right)^2 - 4R \cdot P_{ESS}}}{2R} \tag{33}$$

To estimate the internal temperature of the battery cell, two parts of the battery thermal model are considered: thermal generation and thermal transfer. Based on [29], under large terminal current, the battery heat generation is dominated by the ohmic heat generated over the internal resistance R of the

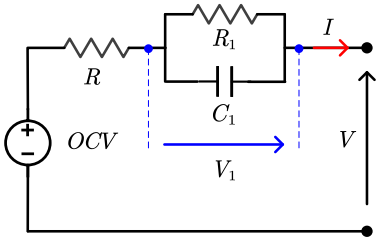

Fig. 4. Battery electric circuit model.

battery cell. The battery heat generation is proportional to the square of the terminal current and thus given as

$$Q = I^2 \cdot R \tag{34}$$

Taking the terminal current (33) into (34), heat generation can be obtained. To build a simplified lump thermal model, it is assumed that the battery shell temperature and internal temperature are both uniform, and heat generation is uniformly distributed within the battery. Heat conduction is assumed to be the only heat transfer form between the battery internal and shell, and between the battery shell and the ambience. Thus the thermal model of the battery cell is given as

$$C_1 \frac{dT_{in}}{dt} = Q - k_1 \cdot (T_{in} - T_{sh}) \tag{35}$$

$$C_2 \frac{dT_{sh}}{dt} = k_1 \cdot (T_{in} - T_{sh}) - k_2 \cdot (T_{sh} - T_{amb}) \tag{36}$$

where T_{in} and T_{sh} are battery internal and shell temperature, respectively; T_{amb} is the ambient temperature; C_1, C_2 are the battery internal and shell thermal capacity, respectively; Q is the generated power in (34); k_1 is the heat conduction coefficient between the battery internal and the shell, and k_2 the heat conduction coefficient between the battery shell and the ambience. Discretizing (35) and (36), the internal and shell temperatures of the battery cell can be expressed as (37) by substituting (33) into (34).

$$\begin{bmatrix} T_{in}(k+1) \\ T_{sh}(k+1) \end{bmatrix} = \begin{bmatrix} 1 - T_s \cdot \frac{k_1}{C_{q1}} & T_s \cdot \frac{k_1}{C_{q1}} \\ T_s \cdot \frac{k_1}{C_{q2}} & 1 - T_s \cdot \frac{k_1 + k_2}{C_{q2}} \end{bmatrix} \cdot \begin{bmatrix} T_{in}(k) \\ T_{sh}(k) \end{bmatrix} + \begin{bmatrix} T_s \cdot \frac{f_R(T_{in}(k)) \cdot I^2(k)}{C_{q1}} \\ T_s \cdot \frac{k_2 \cdot T_{amb}}{C_{q2}} \end{bmatrix} \tag{37}$$

The established model has been validated with high fidelity. As being confirmed in [29], [31], the root-mean-square error (RMSE) of the thermal voltage is around 3.4mV, which is equivalent to 0.1% of the battery nominal voltage. The internal temperature modeling RMSE is $0.37°$C and the maximum error is $1.28°$C. Therefore, the accuracy of the established electrothermal model is acceptable. The electrothermal model can hence be used as a good benchmark for battery packs in real-life applications.

III. Problem formulation

In this paper, the optimization objectives aim to compensate the reactive power of the traction loads by utilizing the PFC to satisfy the power quality requirements, as well as to shave the peak traction power and smooth the regenerative power with the aid of the integrated ESS. Meanwhile, to enhance the system reliability and increase the lifespan of both the power electronics modules and the integrated ESS, thermal constraints are also considered in the optimization.

A Markov Decision Process (MDP) model with a discrete-time step is built. The whole sequential decision-making process of the MDP model can be described as follows: given a set of states s_k at time step k which includes ESS states, power electronics module thermal states, traction load states, and ambient temperature, the agent selects a set of actions from an action-space $a_k \in A$ based on the policy π. The goal of the proposed algorithm is to find the optimal policy to achieve the proposed optimization objectives. The MDP is formulated as:

A. State Space

For the PFC converters (Converters A and B) and DC-DC converters in the co-phase system with ESS integration, the state space at time instant k is defined as $s_{k.Con} = (T_{Con.A}(k), T_{Con.B}(k), T_{Con.DC}(k))$, where $T_{Con.A}(k)$, $T_{Con.B}(k)$ and $T_{Con.DC}(k)$ represent the temperatures of the converter A, converter B in PFC, and DC-DC converter respectively. It is assumed that the heat is uniformly distributed within the converters. Thus the temperature of each submodule in the converters can be regarded as the same. Furthermore, each submodule of the PFC converters and DC-DC converters includes two IGBTs and two diodes as shown in Fig. 1(b). The maximum value of the temperatures of the IGBTs and diodes in each submodule is seen as the temperature of the submodule, also the temperature of the converter, i.e.

$$\begin{cases} T_{Con.A}(k) = \max\left(T_{PA-T1/T2}(k), T_{PA-D1/D2}(k)\right) \\ T_{Con.B}(k) = \max\left(T_{PB-T1/T2}(k), T_{PB-D1/D2}(k)\right) \\ T_{Con.DC}(k) = \max\left(T_{D-T1/T2}(k), T_{D-D1/D2}(k)\right) \end{cases} \tag{38}$$

The thermal states of the converters defined in (38) at each time instant can be obtained by (22)-(24). For the ESS, the state space at time instant k is defined as $s_{k.ESS} = (SoC(k), T_{in}(k), T_{sh}(k))$. It is also assumed that the charging/discharging power and temperature are uniformly distributed within the whole ESS. Thus the states of each battery cell can stand for the states of the whole ESS. The state transition function reflecting the electrothermal dynamics of the ESS from k to $k+1$ is defined in (27) and (37). Besides, for the traction loads, the state of $s_{k.L} = (S_L(k), P_{ave.L}(k))$ includes the apparent traction power and average active traction power. The power factor of the traction loads is set to 0.95 based on [32]. Finally the ambient temperature is defined as $s_{k.amb} = T_{amb}(k)$. The state in the MDP at the time instant k can be expressed as $s_k = (s_{k.Con}, s_{k.ESS}, s_{k.L}, s_{k.amb})$

B. Action Space

As stated in Section II.A, once the phase angle φ and the power of ESS are determined, the power flow can be calculated accordingly. The size of the actions is a hyper-parameter. If the size is too small, the actions are not fine-grained, and the agent may have difficulties to reach a specific location with more accuracy. While if the size is too large, it may slow down the learning process or lead to sub-optimal solutions as it takes more steps to reach a specific targeted location. Generally speaking, the more choices are designed for the space discrete actions, the worse for the algorithm convergence. Based on these considerations, the action space of the ESS is discretized into 9 actions as

$$a_{k.ESS} = (-1, -0.75, ..., 0, ..., 0.75, 1)\, P_{ESS}^{\max} \tag{39}$$

where $P_{ESS}^{\max}$ is the maximum discharging/charging power. The action space of phase angle is defined as

$$a_{k.\varphi} = (-1, -0.8, ..., 0, ..., 0.8, 1)\,\varphi_{\max} \tag{40}$$

where there are 11 actions in total, $\varphi_{\max}$ is the maximum allowable phase angle which also reflects the maximum reactive power feeding into the three-phase power grid. $\varphi_{\max}$ is set to 20° based on the requirements [32]. Thus the whole action space $a_k = (a_{k.ESS}, a_{k.\varphi})$ includes 9×11 actions in total, and the action intervals are set to be 1 MW for the ESS and 2° for the phase angle, which are fine-grained and reasonable according to the findings in [33], [34].

C. Reward

The design of the reward significantly affects the optimization results. To achieve the optimization objectives illustrated above, the reward at time step k is defined as (41).

$$R_k = -\left(\underbrace{a \triangle P(k)}_{load\ shifting} + \underbrace{b\rho_{pa}(k)}_{phase\ angle} + \underbrace{c\rho_T(k)}_{thermal} + \underbrace{d\rho_{SoC}(k)}_{SoC}\right) \tag{41}$$

where a, b, c and d are the weighting factors implying the different importance of the objectives and also enforcing the objectives dimensionally compatible with each other.

In (41), the first part $\triangle P(k)$ considers the peak load shaving. It is defined as

$$\triangle P(k) = |P_{ave.L}(k) - (P_L(k) - P_{ESS}(k))| \tag{42}$$

where $P_{ave.L}(k)$ represents the average power of the traction loads during the timeslot $[0, k]$. When $\triangle P(k)$ is reduced, it means the power from the three-phase power grid is smoother. The second part in (41) is the penalty term for the phase angle, which is defined as (43) to reduce the phase angles so that the lower reactive power is fed into the three-phase power grid.

$$\rho_{pa}(k) = |\varphi(k)| \tag{43}$$

The third part $\rho_T(k)$ in (41) is the thermal penalty term of the converters and ESS, which is defined as

$$\rho_T(k) = \sum \rho_{T(\cdot)}(k) \tag{44}$$

$$\rho_{T(\cdot)}(k) = \begin{cases} 0 & \text{if } T(\cdot) \leqslant T(\cdot)_{\lim} \\ T(\cdot) - T(\cdot)_{\lim} & \text{if } T(\cdot) > T(\cdot)_{\lim} \end{cases} \tag{45}$$

where $(\cdot)$ represents the ESS or converters. $T(\cdot)_{\lim}$ is the temperature limit for the ESS or converters.

In (41), the term $\rho_{SoC}(k)$ is an indicator used for the SoC limit, which is defined as

$$\rho_{SoC}(k) = \begin{cases} 0 & \text{if } SoC^{\min} \leqslant SoC \leqslant SoC^{\max} \\ 1 & \text{otherwise} \end{cases} \tag{46}$$

The objective function, i.e. the cumulative rewards over the whole optimization process is denoted as:

$$\max\ f = \sum_0^K R_k \tag{47}$$

D. Constraints

In the formulated optimization problem, the constraints for the operation of railway co-phase system integrated with ESS are considered below. The power balance constraint can be found in (1) and (2). For the converters and ESS, the thermal limit can be expressed as (48) and considered as a soft constraint in (47). The power limit (49) of the ESS is imposed in the action space, and the SoC limit of the ESS can be expressed as (50) and added in (47) as the hard constraints by setting a rather large weighting factor.

$$T(\cdot) \leqslant T(\cdot)_{\lim} \tag{48}$$

$$0 \leqslant |P_{ESS}| \leqslant |P_{ESS}^{\max}| \tag{49}$$

$$SoC^{\min} \leqslant SoC \leqslant SoC^{\max} \tag{50}$$

IV. Methodology

In this section, based on the formulated MDP, a DRL-based thermal constrained energy management strategy is proposed to solve the formulated problem in Section III. Moreover, an FRA-based pruning method is developed to reshape the DRL agent and make the agent more compact without sacrificing its performance.

A. Prioritized DQN

The deep Q-learning network (DQN) originated from the original Q-learning algorithm is one of the state-of-the-art DRL algorithms. Q-learning is a model-free RL algorithm aiming at training an agent to find the best policy with a set of given states by exploring the environment [25]. The action taken under a specific set of given states is determined by the action-value Q function which is denoted by $Q_\pi(s, a)$ defined as

$$Q_\pi(s, a) = \mathbb{E}_\pi\left[\sum_0^K \gamma^k \cdot R_k \,|s_k = s, a_k = a\right] \tag{51}$$

where γ is the discount factor. The policy π maps from the current states to the specific actions.

To find the best policy π^*, the Q function $Q_\pi(s, a)$ needs to be literately updated by interacting with the environment based on the following Bellman equation:

$$\begin{aligned} Q(s_k, a_k) \leftarrow\ & Q(s_k, a_k) \\ & + \alpha\left[\left(R_k + \gamma \max_a Q(s_{k+1}, a_{k+1})\right) - Q(s_k, a_k)\right] \end{aligned} \tag{52}$$

where α is the learning rate. $Q(s, a)$ is approximated by a look-up table characterized by a set of states and actions. When the iteration reaches the preset value, it can be considered that $Q(s, a)$ converges to the approximate optimal Q-value $Q_\pi^*(s, a)$. Then the action will be taken based on the most updated $Q(s, a)$ by using ϵ-greedy strategy. The rule is the action $a_k = arg\max_a Q(s, a)$ is taken with the probability 1-ϵ or a random action is selected with the probability ϵ [35].

However, Q-learning will face a serious challenge, that is when the state and action space is high-dimensional, the built look-up table is rather large. To address this challenge, an

algorithm proposed by Deepmind [35] is to use DNN to approximate the optimal Q-value function $Q_{\pi}^{*}(s,a)$ rather than a lookup table. The approximated optimal action-value Q function by the DNN can be denoted by

$$Q(s,a,\omega) \approx Q_{\pi}^{*}(s,a) \tag{53}$$

where ω is the weights of the neural network. During the training process of DQN, the weights are updated by minimizing the mean square error (MSE) loss between the values of Q target and Q predict:

$$L(\omega) = \mathbb{E}[(\overbrace{R_k + \gamma \max_{a_{k+1}} Q(s_{k+1}, a_{k+1}, \omega^{-})}^{Q\ target} - \overbrace{Q(s,a,\omega)}^{Q\ predict})^2] \tag{54}$$

The Q target value is $y_k = R_k + \gamma \max_{a_{k+1}} Q(s_{k+1}, a_{k+1}, \omega^{-})$, where $Q(s_{k+1}, a_{k+1}, \omega^{-})$ is a separate neural network namely target network. The weights ω^{-} in the target network are updated periodically by copying the weights ω in the predict network $Q(s,a,\omega)$ to replace the previous weights ω^{-}.

In the DQN training process, the recent experiences (s_k, a_k, R_k, s_{k+1}) are stored in the experience replay pool. In each training step, a batch of samples are uniformly selected from the experience replay pool for the DNN-based agent training based on (54). However, due to the limit of the size for the experience replay pool, the memory only stores recent M samples by removing the oldest experiences for allocating a space for the latest ones. However, it is highly likely that the experiences with higher impact on the learning are forgot during the above process. To resolve this challenge, a new technique prioritizing significant experiences is proposed in [36] by changing the sampling distribution of DQN training. The overall idea for prioritized experience replay is that the samples with higher TD error obtains higher ranking in terms of probability than the other samples. Then in the training process, the experiences are stochastically sampled based on the assigned probabilities. In the case studies presented in this paper, such prioritized DQN is adopted.

B. FRA-based Pruning Method for DQN Agent

In the field of system identification, a structure, namely linear-in-the-parameter model, is widely adopted for approximating a large class of nonlinear systems. Such nonlinear models form a linear combination of model terms, or basis functions, which are nonlinear functions of the system variables. However, one problem with such models is that an excessive number candidate model terms or basis functions have to be considered, which may lead to high computational complexity or over-fitting. To handle this issue, the model selection algorithms have been proposed to build parsimonious and compact models with a much smaller number of model terms. Among theses algorithms, a particular one, namely fast recursive algorithm (FRA), was proposed to fast select the most significant model terms as well as identify the model parameters [37]. The FRA is introduced below. Consider a nonlinear discrete-time dynamic system denoted by a linear-in-the-parameter model as (55), which is identified by N data samples $\{x(i), y(i)\}_{i=i}^{N}$

$$y = \Psi\Theta + \Xi \tag{55}$$

where $y = [y(1), ..., y(N)]^T \in \Re^N$ represents the system outputs. $\Psi = [\varphi_1, ..., \varphi_j, ..., \varphi_S] \in \Re^{N\times S}$ is the regression matrix containing all candidate model terms. In the regression matrix, each element $\varphi_j = [\varphi_j(x(1)), ..., \varphi_j(x(N))]^T \in \Re^{N\times 1}$ $(j = 1, ..., S)$ represents a specific nonlinear function of the N inputs. $\Theta = [\theta_1, ..., \theta_S]^T \in \Re^S$ is the parameter matrix including S unknown system parameters to be identified. $\Xi = [\varepsilon(1), ..., \varepsilon(N)]^T \in \Re^N$ is the system residual matrix.

Before introducing the principle of FRA, two recursive matrices, M_k and R_k, are predefined as (56) and (57) to facilitate the further demonstration of the fast model term selection and parameter identification.

$$M_k = \Psi_k^T \Psi_k \tag{56}$$

$$R_k = 1 - \Psi_k M_k^{-1} \Psi_k^T \tag{57}$$

where $\Psi_k \in \Re^{N\times k}$ $(k = 1, ..., S)$ is the submatrix containing the first k columns of the regression matrix Ψ. Notably, it is defined that $R_0 = 1$.

To identify the system parameters, a cost function as (58) can be defined

$$\min E(\Theta) = (\Psi\Theta - y)^T (\Psi\Theta - y) \tag{58}$$

Generally, when the first k columns in matrix Ψ are selected, the parameter matrix $\hat{\Theta}_k \in \Re^k$ that minimizes the cost function (58) can be estimated as (59) by the least square method if Ψ_k is of full column rank.

$$\hat{\Theta}_k = M_k^{-1} \Psi_k^T y \tag{59}$$

The minimal cost function associated with $\hat{\Theta}_k$ can be reformulated as

$$E_k = y^T y - \hat{\Theta}_k^T \Psi_k^T y = y^T R_k y \tag{60}$$

Based on (60) and the properties of R_k given in [37], it follows that

$$\begin{cases} E_{k+1} = y^T R_{k+1} y = E_k - \frac{y^T R_k \varphi_{k+1} \varphi_{k+1}^T R_k^T y}{\varphi_{k+1}^T R_k \varphi_{k+1}} \\ E_0 = y^T y \end{cases} \tag{61}$$

Furthermore, to simplify the computational complexity, three quantities are defined as follows

$$\begin{cases} \varphi_j^{(k)} \triangleq R_k \varphi_j, \varphi_j^{(0)} \triangleq R_0 \varphi_j = \varphi_j \\ a_{k,j} \triangleq \left(\varphi_k^{(k-1)}\right)^T \varphi_j^{(k-1)}, a_{1,j} \triangleq \varphi_1^T \varphi_j \\ b_k \triangleq \left(\varphi_k^{(k-1)}\right)^T y, b_1 \triangleq \left(\varphi_1^{(0)}\right)^T y = \varphi_1^T y \end{cases} \tag{62}$$

where $j = 1, ..., S$ and $k = 0, 1, ..., S$. According to (62), the net contribution $\triangle E_{k+1}$ of the new model term φ_{k+1} to the

cost function when it is included in the model can be expressed by

$$\triangle E_{k+1}=\frac{\left(y^T\varphi_{k+1}^{(k)}\right)^2}{\left(\left(\varphi_{k+1}^{(k)}\right)^T\varphi_{k+1}^{(k)}\right)}=\frac{\left(b_{k+1}^T\right)^2}{a_{k+1,k+1}},k=0,...,S-1 \tag{63}$$

By estimating the net contribution of each term using (63), the importance of all model terms can be ranked and a total of k model terms with maximum net contributions are selected one by one. After that, the parameters of the k selected model terms can be calculated by

$$\hat{\theta}_j=\frac{b_j-\sum_{i=j+1}^{k}\hat{\theta}_i a_{j,i}}{a_{j,j}},j=k,k-1,...,1 \tag{64}$$

Notably, $\sum_{i=k+1}^{k}\hat{\theta}_i a_{k,i}=0$.

As aforementioned, a deep neural network, as the agent in the DQN algorithm, is used to approximate the action-value Q function. Generally, the adopted deep neural network for the DQN agent has multiple layers and neurons, which results in a large amount of model parameters. It will cost large computation and memory resources so as to limit the implementation of the trained agent in the real energy flow controller. To reduce the complexity of the agent, an FRA-based network pruning method is proposed to reshape the trained agent of DQN by removing the redundant neurons based on the net contributions of all neurons. Meanwhile, based on the proposed FRA-agent pruning method, the model parameters (neuron wights) can be re-assigned for the pruned agent to guarantee the accuracy of the approximation of the action-value Q function.

The adopted DNN for the DQN agent in this paper has four layers, including one input layer, one output layer and two fully-connected (FC) hidden layers. Suppose $X_0\in\Re^{N\times L}$ is the input matrix to the input layer, where N is the number of the samples stored in the replay memory of DQN and L represents the number of the states in each sample, i.e. $L=num\left(S\right)$. For the input layer, the outputs are equal to the inputs, i.e. $Y_0=X_0$. $X_i\in\Re^{N\times(I(i)+1)}$ $(i=1,2,3)$ represents the input matrix to the first hidden layer, second hidden layer and output layer. X_i is constituted by the outputs of the previous layer and a $N\times1$ bias vector with all elements being set to 1, i.e. $X_i=\left[Y_{i-1}\left|[1]_{N\times1}\right.\right]$. $I\left(i\right)$ is the number of the inputs of layer i, which is also equals to the number of neurons of the previous layer, i.e. $I\left(i\right)=num\left(Y_{i-1}\right)$. $Y_i=\left[y_{i,1},y_{i,2},...,y_{i,N}\right]^T\in\Re^{N\times O(i)}$ $(i=1,2,3)$ denotes the output matrix from layer i.$O\left(i\right)$ is the number of the outputs of layer i. Notably, based on the DQN algorithm, for the output layer, $O\left(3\right)$ is equal to the number of actions, i.e. $O\left(3\right)=num\left(A\right)$. The neural network model thus can be denoted as

$$\begin{cases}Y_0=X_0\\X_i=\left[Y_{i-1}\left|[1]_{N\times1}\right.\right]\ \left(i=1,2,3\right)\\\mathcal{Z}_i=X_i\Theta_i+\Xi_i\ \left(i=1,2,3\right)\\Y_i=f\left(\mathcal{Z}_i\right)\ \left(i=1,2,3\right)\end{cases} \tag{65}$$

where $\Theta_i\in\Re^{(I(i)+1)\times O(i)}$ represents the parameter matrix of layer i, which consists of the $I\left(i\right)\times O\left(i\right)$ weight matrix and $1\times O\left(i\right)$ bias vector. $\mathcal{Z}_i$ is an intermediate matrix representing the linear relationship of the inputs. Ξ_i is the residual vector of layer i. $f\left(\cdot\right)$ represents the activation function.

To remove the redundant neurons of the trained agent without sacrificing the overall performance, take layer i as an example. The detailed procedure including four steps for the neuron selection and network reshaping is illustrated as follows:

Step 1: Calculate the net contribution of each input in layer i. Suppose $\mathcal{Z}_i=\left[z_{i,1},z_{i,2},...,z_{i,N}\right]^T\in\Re^{N\times O(i)}$ is the intermediate matrix of layer i. Based on (63), the net contribution of each input in layer i can be calculated by

$$\triangle E_{k+1}=\frac{\left(\left(\mathcal{Z}_i\right)^T x_{k+1}^{(k)}\right)^2}{\left(\left(x_{k+1}^{(k)}\right)^T x_{k+1}^{(k)}\right)},k=0,1,...,I\left(i\right)-1 \tag{66}$$

Notably, the net contribution of each input in layer i also reflects the importance of each neuron of layer $i-1$. Then rank the inputs based on their net contributions

Step 2: Select the highest-ranked M inputs with M initially being set to 1, and then re-assign the parameters for the selected inputs based on (67)

$$\hat{\theta}_{i,j}=\frac{\left(x_j^{(j-1)}\right)^T\mathcal{Z}_i-\sum_{p=j+1}^{k}\hat{\theta}_p\left(x_j^{(j-1)}\right)^T x_p^{(j-1)}}{\left(x_j^{(j-1)}\right)^T x_j^{(j-1)}},\quad j=M,M-1,...,1 \tag{67}$$

where M is the total number of the selected inputs. Define $\hat{\Theta}_i=\left[\hat{\theta}_{i,1},\hat{\theta}_{i,2},...,\hat{\theta}_{i,M}\right]^T$

Step 3: Evaluate the performance of the reshaped layer i. The outputs of the reshaped layer i can be calculated as

$$\begin{cases}\hat{\mathcal{Z}}_i=\hat{X}_i\hat{\Theta}_i+\Xi_i\\\hat{Y}_i=f\left(\hat{\mathcal{Z}}_i\right)\end{cases} \tag{68}$$

The root mean square error between $\hat{Y}_i$ and Y_i over N samples is given as

$$\begin{aligned}RMSE&=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\left(\hat{y}_{i,n}-y_{i,n}\right)\left(\hat{y}_{i,n}-y_{i,n}\right)^T\right)}\\&=\frac{1}{\sqrt{N}}\left\|\hat{Y}_i-Y_i\right\|\end{aligned} \tag{69}$$

where $\|\cdot\|$ is the Euclidean norm.

Step 4: Repeat steps 2-3 until M reaches the maximum number of inputs in layer i with M increased by 1 in each cycle, or the value of RMSE in (69) is smaller than the predefined limit.

Based on the Steps 1-4, the flowchart of the FRA-based pruning method for the DQN agent is demonstrated in Fig. 5.

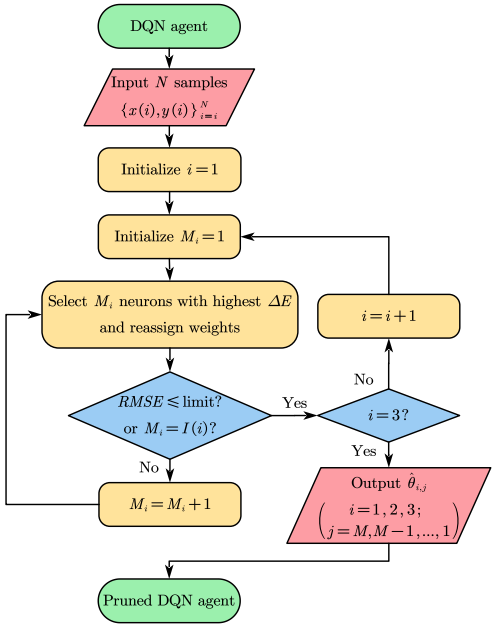

Fig. 5. Flowchart of FRA-based agent pruning method.

C. FRA-pruned DQN-Based Thermal Constrained Energy Management

To solve the formulated optimization problem in (47) under the DQN-based framework, the models of the thermal-conscious railway co-phase system integrated with ESS are built. With the defined system states, actions and rewards in Section III, the flowchart of the DQN-based thermal constrained energy management is shown in Fig. 6. The agent is trained by interacting with the environment. With a set of given actions, the transition of the system states can be obtained by the models introduced in Section II. Meanwhile the reward corresponding to this set of actions is also obtained. With the training process going forward, the agent will gradually converge guided by the reward function. After the training, the proposed FRA-based pruning method is adopted to make the DQN agent more compact. It is worth noting that the DNN parameters of the agent are fixed after pruning. And the pruned agent deployed in the controller can directly give a set of optimal actions for the energy management based on the real-time system states.

V. Case Studies

In this section, to verify the effectiveness of the proposed FRA-pruned DQN based thermal constrained energy management strategy, comprehensive case studies are performed on a PC with a 2.80GHz Intel i7-7700HQ CPU and 16GB RAM. The algorithm is developed on Python 3.0 and Keras.

A. Case Description and Input Parameters

Taking a high-speed railway station in China as an example, the traction load curve with a sampling time of 15 mins is shown in Fig. 7 [38], where the positive values represent the power fed into the trains, and the negative ones represent the regenerative power from the braking trains. The whole optimization time is one day including 96 intervals (15 mins per interval). The time period is set to 1 second for the precise system state iteration. The ambient temperatures of the whole day are shown in Fig. 7.

The management strategy is highly related to the weighting factors selected in (47). Generally speaking, a larger weighting factor suggests greater importance of the objective or more stringent compliance to the imposed constraints, while a smaller weighting factor implies less importance of the objectives or softer constraints allowing certain level of violation tolerance of the constraints. The weighting factors are therefore set as the trade-off between different objectives and constraints, and in this paper, a, b, c and d are set as 5, 1, 10 and 1000, respectively. Among them, the SoC constraint is seen as a hard constraint that has to be strictly observed under any circumstances, thus a rather large value of 1000 is assigned for d. Meanwhile, the temperature is desired to be controlled below a certain value to avoid unexpected degradation or thermal runaway incidents. However, such constraints may be violated in practice due to the uncertainties stemmed from estimation errors and external disturbances. Instead of using a hard constraint, therefore, the thermal constraint is set as a soft constraint allowing certain level of violation tolerance. The penalty term c is assigned with a relatively large value of 10. Unlike the above constraints, the objectives of peak power shaving and phase angle are not directly linked to the system safety and reliability, and both are considered to be part of the system operation optimization problem. Thus the weighting factors a and b are assigned with a relatively small value of 5 and 1, respectively. The objective of the phase angle is assigned with a less important factor than that of the peak power shaving since the actions for the phase angle have been limited in an allowable range. That implies that even if the phase angle is not considered in (47), the phase angle will still satisfy the operation requirement of the power grid. Furthermore, the settings of the weighting factors have already taken into account the dimensional compatibility issue of different objectives. It is also worth noting that the trained agent under a specific set of weighting factors can provide a representative solution of the whole Pareto front in a multi-objective optimization problem, which can well be used to verify the performance of the proposed method.

In the co-phase system, the ESS integrated PFC has 15 SMs in each arm, 120 SMs in total for eight arms. The design parameters of the PFC and DC-DC converters are listed in Table I. The SEMIKRON IGBT module SKiiP 1213 GB123-2DL V3 [39] is used for the PFC converters, and SEMIKRON IGBT module SEMiX453 GB12E4p [40] is used for the DC-DC converters. The parameters of the selected power electronics modules are listed in Table II. Based on the datasheets [39], [40], the operation temperature of power electronics modules is normally within the range of $-40°$C to $85°$C. The lifetime will be significantly reduced when the junction temperature is higher than $85°$C. Thus the penalty term of the overtemperature for the power electronics modules is triggered from $85°$C. For the ESS, a LiFePO4 Graphite battery is used in this paper. One battery cell has a capacity

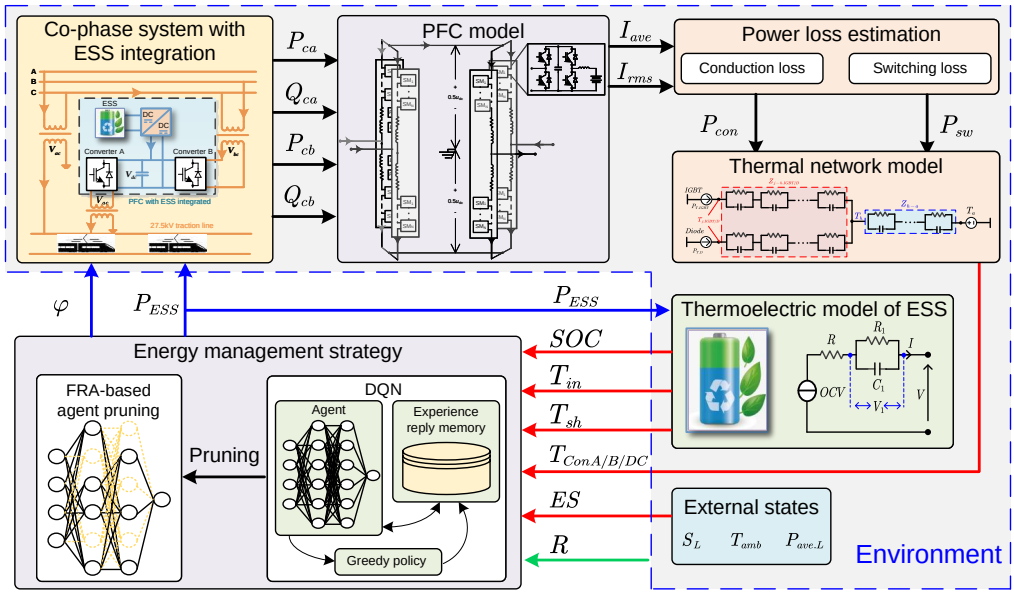

Fig. 6. Flowchart of FRA-pruned DQN-based thermal constrained energy management strategy.

TABLE I
PARAMETERS OF PFC AND DC-DC CONVERTERS

Parameters of PFC	Values	Parameters of DC-DC converter	Values
Number of SMs in each arm	15	Rated voltage of battery side	832 V
Rated voltage of AC side	10 kV	Rated voltage of PFC side	1000 V
Rated DC-link voltage of each SM	1000 V	Maximum current ripple	0.3
Switching frequency	1000 Hz	Switching frequency	10 kHz

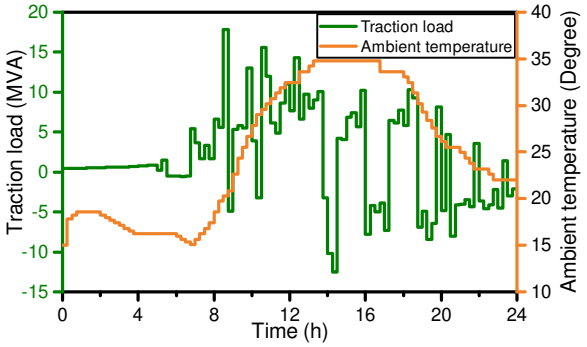

Fig. 7. Traction loads and ambient temperature.

TABLE II
PARAMETERS OF POWER ELECTRONICS MODULES

Modules SKiiP 1213 GB123-2DL V3/ SEMiX 453 GB12E-4p			
Parameters	Values	Parameters	Values
$E_{on/off}$/ mJ	390/ 45	$K_{v.T}$	1.3/ 1.3
E_{rr}/ mJ	56/ 30	$K_{v.D}$	0.6/ 0.6
$C_{E.T}$	0.003/ 0.003	I_{ref} /A	600/ 450
$C_{E.D}$	0.006/ 0.006	V_{ref} /V	900/ 600
$K_{i.T}$	1/ 1	T_{ref} /°C	125/ 150
$K_{i.D}$	0.6/ 0.6		

of 10Ah and a nominal operating voltage of 3.2V. Based on [29], the internal resistance will rise noticeably when the SoC falls below 20%, which would lead to a large internal heat generation and speed up the battery degradation. Thus battery is only cycled in the range higher than 20% SoC. Meanwhile, within this range, the battery internal resistance does not change much with SoC variation, and the accuracy of the adopted ESS electrothermal model will then become relatively high as the internal resistance is primarily related to the internal temperature within this SoC range. The parameters of the ESS and the battery cell are listed in Table III. Based on [29], in real applications, the lithium-ion battery would be degraded faster when its internal temperature is greater than $40°$C. This has also been confirmed in [10], [18], that the lifespan of the ESS would be reduced quickly when the working temperature is higher than $40°$C. Given the aforementioned considerations, the penalty-triggering temperature in the objective function is set to $40°$C. Furthermore, the soft constraint for the overtemperature is also set at $40°$C. The hyperparameters used in the DQN training are listed in Table IV, their settings are experience-based from the literature [10], [18], [25], [33], [41].

To verify the proposed FRA-pruned DQN-based thermal constrained energy management, five scenarios are investigated:

1) Rule-based energy management under phase angle $\varphi = 0°$ without ESS integration
2) Rule-based energy management under phase angle $\varphi = 20°$ without ESS integration
3) DQN based thermal constrained energy management without ESS integration

TABLE III
PARAMETERS OF ESS

Parameters	Values	Parameters	Values
Rated capacity of ESS	5 MWh	a_1	0.982
Maximum power of ESS	5 MW	b_1	2.1e-4
Range of SoC	0.2-1.0	c_1	264.7
Rated voltage of battery cell	3.2 V	c_2	30.7
Rated capacity of battery cell	10 Ah	k_1	1.286
		k_2	0.3009
OCV		See [29]	
R		See [29]	

TABLE IV
HYPERPARAMETERS USED FOR TRAINING THE AGENT

Parameters	Values	Parameters	Values
No. of hidden layers	2	Total episodes	10000
No. of neurons	80 / 250	Learning rate	0.001
Activation function	ReLU	Minibatch size	64
Experience pool size	20000	Discount factor	0.98

4) DQN based energy management with ESS integration
5) DQN based thermal constrained energy management with ESS integration

The scenarios examined to verify the performance of the proposed DQN-based thermal-constraint energy management strategy cover the optimization of various control objectives for the railway co-phase system integrated with or without the ESS, and the thermal performance of both the power electronics devices and ESS are examined. The first three scenarios presented in Section V.B consider the railway co-phase system without ESS. The purpose is to verify and compare the performance of the proposed strategy against popular rule-based strategies in reactive power compensation while satisfying the thermal constraints. While the next two scenarios presented in Section V.C consider the railway co-phase system integrated with ESS, and the purpose is to verify the performance of the proposed DQN-based thermal-constraint energy management strategy in compensating the reactive power, shaving the peak load while meeting the thermal constraints for both the power electronics modules and ESS. Finally, in Section V.D, the performance of the FRA-based pruning method in reducing the agent complexity is verified.

B. Performance of DQN-based Strategy for Thermal Management without ESS Integration

In this case, the performance of the DQN-based strategy for thermal management without ESS integration is verified. The PFC is only used to compensate the reactive traction loads. In [42], a full compensation strategy with $\varphi = 0^\circ$ is applied to a 10MVA-rated co-phase system in China. $\varphi = 0^\circ$ implies that the three-phase power grid is fully balanced and all reactive power of the traction loads is compensated by the PFC. This full compensation strategy is adopted first in scenario 1 for the comparison. Under the traction loads as shown in Fig. 7, the temperatures of converters A and B over the whole day are shown in Fig. 8(a). The maximum temperatures of converters A and B are 113.7°C and 89.9°C respectively. Besides, another partial compensation strategy discussed in [43] is also applied in scenario 2 for the comparison purpose. The phase angle is set to $\varphi = 20^\circ$ since $\varphi = 20^\circ$ is the maximum allowable phase angle for the three-phase power grid. The temperatures of converters A and B are shown in Fig. 8(b). The maximum temperatures of both converters are 74.8°C and 92.5°C respectively.

When the DQN-based thermal management strategy is adopted in scenario 3, the temperatures of converters A and B at different time instants are shown in Fig. 8(c). The maximum temperatures of converters A and B are 85.9 °C and 91.5 °C respectively. Considering both converters A and B, the maximum temperatures under the rule-based energy management strategies with $\varphi = 0^\circ$, $\varphi = 20^\circ$ and under the DQN-based strategy all exceed the limit by 28.7°C, 7.5°C, and 6.5°C respectively. While it is clear that DQN-based strategy performs better for constraining the thermal performance. To further verify the performance of the proposed DQN-based thermal management strategy, an index reflecting the cumulative values of temperature exceeding the limit is defined as:

$$CT(k) = \sum_{0}^{k} \rho_{T.(\cdot)}\left(k\right) \tag{70}$$

where the definition of $\rho_{T(\cdot)}\left(k\right)$ has been given in (45). $CT(k)$ of converters A and B over the whole day is shown in Fig. 9, where $CT(k)$ of PFC is defined as the sum of $\rho_{T(\cdot)}\left(k\right)$ of converters A and B at each time instant. It can be seen that, at the final time instant of the whole day, $CT(k)$ of PFC based on the proposed DQN-based strategy is the smallest, which is shown as the solid blue line in Fig. 9. The partial compensation strategy with $\varphi = 20^\circ$ has a similar but slightly larger $CT(k)$ value (solid green line) than the proposed strategy. The full compensation strategy with $\varphi = 0^\circ$ has the largest $CT(k)$ value (solid red line). It implies that, over the whole day, the temperatures of the PFC are overall below the limit under the DQN-based strategy. Even if the temperatures exceed the limit at some specific time instants, they do not stretch the limit significantly and the undesirable thermal condition does not last long either. The details of $CT(k)$ at the final time instant under different strategies are listed in Table V. It is evident that the proposed DQN-based strategy outperforms the fixed rule-based full and partial compensation strategies in terms of the thermal, and a reduction of 89.3% in the CT values, i.e. the total thermal violation against limits, under Scenario 3 can be achieved by the proposed DQN-based strategy compared to the full compensation strategy (Scenario 1).

It is worth noting that in this case, the thermal management is achieved by optimising the phase angle φ of the three-phase power grid. In Scenario 3, under the proposed energy management strategy, the phase angles are kept at 0° for most of the time, except for 16°, 12°, 8° and 12° at the $35th$, $43rd$, $50th$ and $58th$ time intervals. These four time intervals are exactly the periods when the temperature exceeds the boundary in Scenario 1 which does not consider the thermal constraint. The trained DQN agent is able to respond quickly during the intervals when the thermal constraints are violated to constrain the temperature increase by adjusting the phase

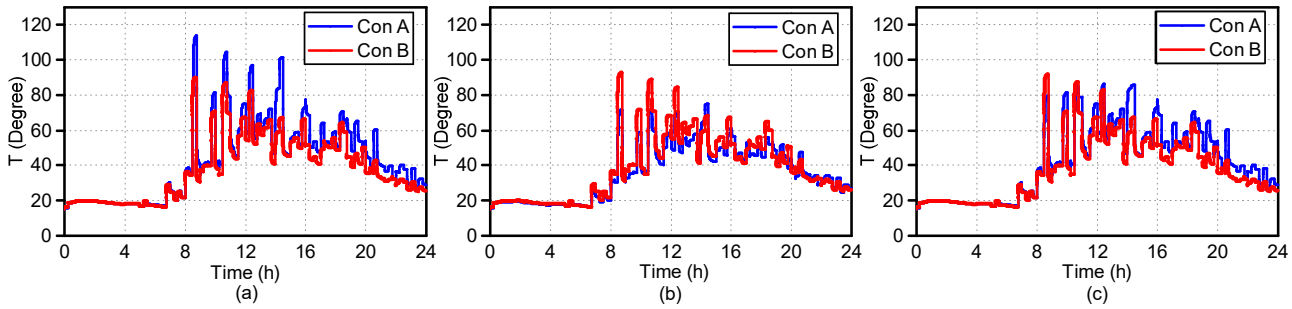

Fig. 8. Thermal performance comparisons of converters without ESS integration. (a) S1; (b) S2; (c) S3 DQN-based strategy.

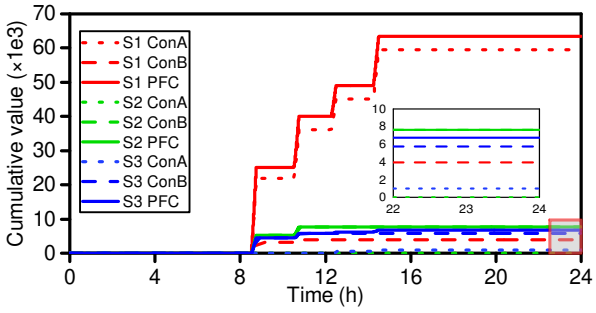

Fig. 9. $CT(k)$ of S1, S2 and S3.

angles. Thus the phase angles do not change violently. When φ is not equal to zero, some reactive power will be injected into the three-phase power grid. The cumulative reactive energy (CRE) injected into the three-phase power grid during the whole day under different strategies are listed in Table V. It is obvious that the reactive power remains zero under the full compensation strategy. While the thermal performance under this strategy is undesirable and even not acceptable. Under the partial compensation strategy with $\varphi = 20°$, the total reactive energy reaches 40.1MVArh during the whole day with the equivalent reactive power (ERP) of 1.7MVAr. Under the DQN-based strategy, the total reactive energy is only 3.6MVArh with a reduction of 36.5MVArh compared to the partial compensation scheme and a relatively small average equivalent reactive power of 0.15MVAr is required for the railway power supply. The amount of reactive power under $\varphi = 20°$ is seen as the maximum allowable reactive power injected to the three phase power grid. Under the proposed DQN-based energy management, the injected reactive power only accounts for 9.0% of the maximum allowable amount. In summary, the proposed DQN-based strategy can not only manage the thermal performance effectively but also reduce the reactive power injected into the three-phase power grid as much as possible.

C. Performance of DQN-based Strategy for Peak Load Shaving and Thermal Management with ESS Integration

In this case, the ESS is integrated into the co-phase system to achieve a more flexible power flow. Two different strategies are applied in this case. Both strategies aim to shave the peak traction power as well as compensate the reactive power by using the DQN-based strategies, but one considers thermal management (Scenario 5) and the other does not (Scenario 4). Fig. 10 (a) shows the required power from the three-phase power grid under both strategies. It is evident that, compared to the original traction power, the maximum required power under both strategies is significantly reduced from 16.9MW to 13.6MW with a 19.5% reduction whether the thermal management is considered or not. The maximum regenerative power is reduced from 11.9MW to 6.9MW with a reduction of 42.0% when the thermal management is not considered, and to 8.1MW with a reduction of 31.9% when considering thermal management. Fig. 10 (b) shows the actions and corresponding SoC of the ESS. The actions of the ESS determined by the strategies without/ with considering thermal management are denoted by the blue and red bars respectively. The bars with positive values represent the discharging actions and the bars with negative values represent the charging actions. It can be seen that the trends of SoC under both strategies are similar. When the SoC and traction power are relatively high, the ESS tends to discharge to shave the peak power. When the SoC is relatively low and regenerative power is high, the ESS is operated in the charging mode to shave the peak regenerative power and also absorb energy for the further peak traction load shaving. While for the strategy taking into account the thermal management, the change of the SoC is much smoother than that without thermal management due to the thermal limitations for the ESS and PFC.

To further demonstrate the effectiveness of the DQN-based strategy for the peak load shaving, a statistical index $\triangle P$ given in (42) reflecting the absolute difference between the average traction power and required power from the three-phase power grid is shown in Fig. 11. The smaller values of $\triangle P$ means that the three-phase power curve is smoother. As shown in Fig. 11, the values of $\triangle P$ at different time instants over the whole day (86400 time instants in total) are sorted from the largest to smallest. The values of $\triangle P$ under both strategies whether considering or not considering thermal management are overall lower than that of the original traction loads. The performance of both strategies for load shaving is similar according to Fig. 11. Also, the same conclusion can be drawn from Table VI since the average difference of $\triangle P$ under both strategies is almost similar, i.e. 3.4MW for Scenario 4 and 3.3MW for Scenario 5 respectively. The details for the peak load shaving under both strategies are listed in Table VI. Even if considering

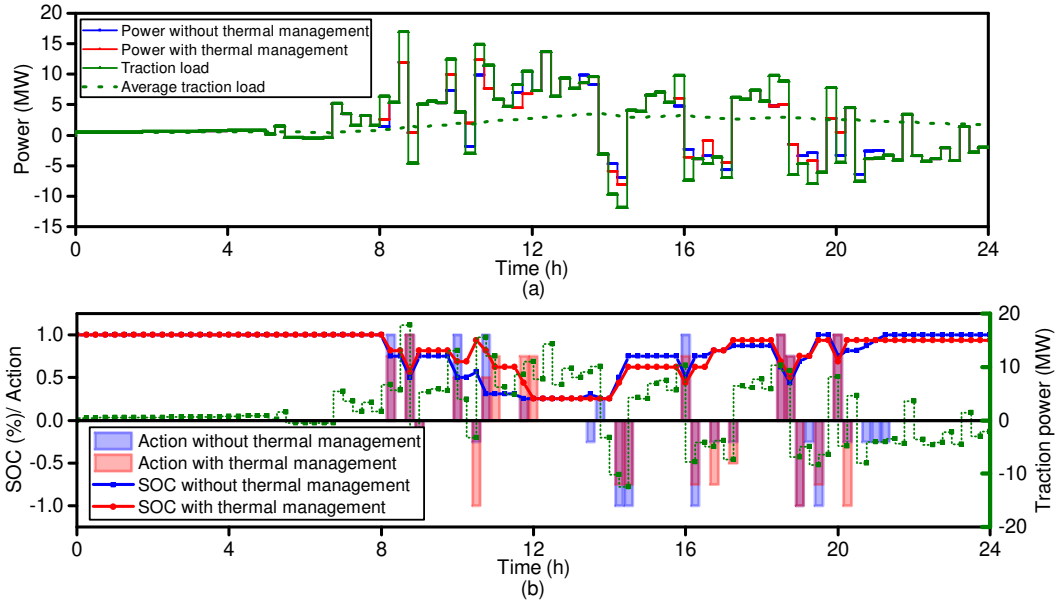

Fig. 10. Result comparisons. (a) Performance comparisons of load shaving; (b) Comparisons of SoC and actions of ESS.

the thermal constraints, the DQN-based energy management strategy can achieve a maximum peak load reduction of 31.9% compared to the original traction loads.

Further, the thermal management performance under both strategies is also analysed. As shown in Fig. 12, with the thermal management, the temperature of ESS is lower than the limit during the whole day. While the maximum temperature of ESS under the strategy without thermal management exceeds the limit, reaching 44.6 °C. For the PFC, compared to the strategy without thermal management, the proposed strategy with thermal consideration performs better in terms of the thermal management. As shown in Fig. 13, the maximum temperatures of converter A under two strategies with and without thermal management are 89.6°C and 124.1°C respectively. The index $CT(k)$ of converter A at the final time instant is listed in Table V. A thermal reduction of 94.1% can be achieved under thermal-constrained strategy (Scenario 5) compared to the strategy without considering the thermal constraint (Scenario 4). The temperatures of converter B and DC-DC converter are below the limit across the whole day under both strategies. Besides, the proposed DQN-based strategy with thermal management only injects 3.4MVArh reactive energy, equivalent to 0.14MVAr reactive power which is acceptable for the railway power supply system. Overall, the proposed thermal constrained DQN-based strategy not only performs well in load shaving, the thermal performance is also much better than the strategy without considering the thermal constraint.

D. Performance of Training Process and FRA-based Pruning for Complexity Reduction

To further verify the superiority of the adopted prioritized DQN algorithm and FRA-based pruning method, comparative studies are conducted. Firstly, as aforementioned, the prioritized DQN algorithm is adopted during the agent training for

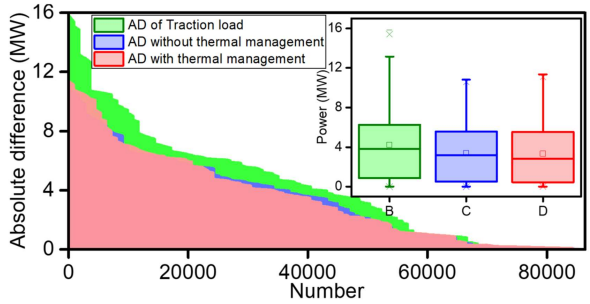

Fig. 11. Comparisons of absolute difference under different strategies.

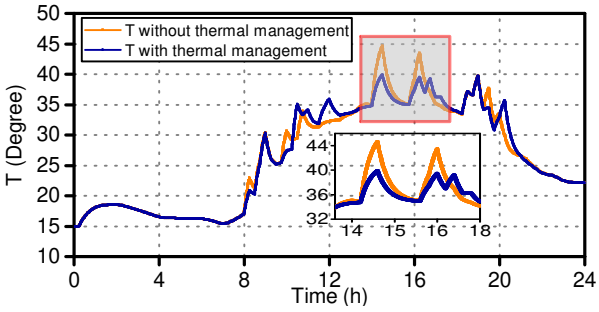

Fig. 12. Thermal performance comparisons of ESS.

fast convergence. Compared to the original DQN proposed in [35], the convergent speed of the prioritized DQN algorithm is theoretically faster. To verify this, the training performance of the original DQN algorithm and the prioritized DQN algorithm used in Scenario 5 are compared in Fig. 14, where the rewards are normalized. In Fig. 14, the solid blue line and orange line represent the average normalized rewards of the prioritized DQN and original DQN per 100 episodes. It can be seen that, as expected, the training performance of the prioritized DQN is better than that of the original DQN since the prioritized

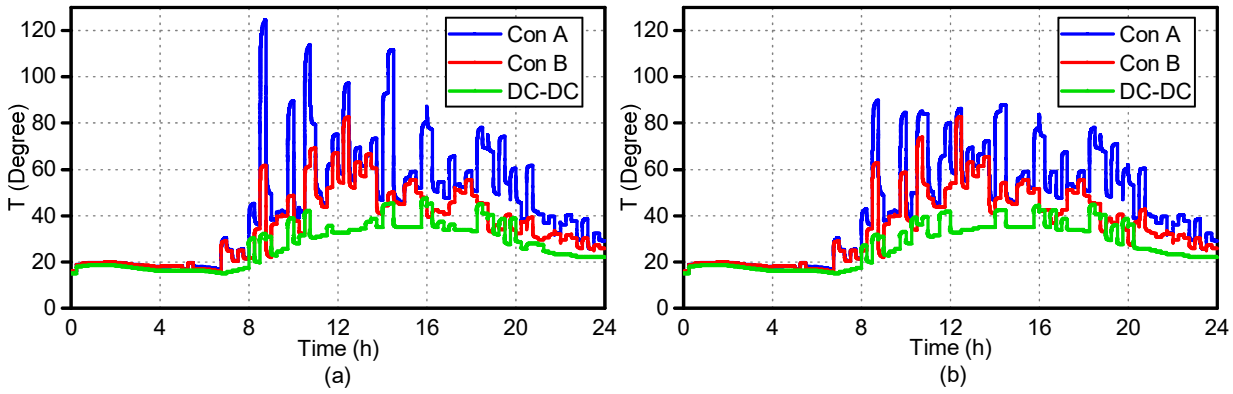

Fig. 13. Thermal performance comparisons of converters. (a) S4 without thermal management; (b) S5 with thermal management.

TABLE V
COMPARISONS OF THERMAL MANAGEMENT AND REACTIVE POWER COMPENSATION

	CT (1e3)				CRE	ERP
	Con A	Con B	PFC	ESS	(MVArh)	(MVAr)
S1	59.6	4.0	63.6	/	0	0
S2	0	7.6	7.6 (88.1% ↓)	/	40.1	1.67
S3	1.0	5.8	**6.8** (**89.3**% ↓)	/	3.6	0.15
S4	93.7	0	93.7	7.2	0	0
S5	5.5	0	**5.5** (**94.1**% ↓)	**0** (**100**% ↓)	3.4	0.14

TABLE VI
COMPARISONS OF PEAK LOAD SHAVING

	Max required power	Max regenerative power	Average difference
Traction load	16.9 MW	11.9 MW	4.2 MW
DQN without thermal management	13.6 MW (19.5% ↓)	6.9 MW (42.0% ↓)	3.4 MW (19.1% ↓)
DQN with thermal management	13.6 MW (19.5% ↓)	8.1 MW (31.9% ↓)	3.3 MW (21.4% ↓)

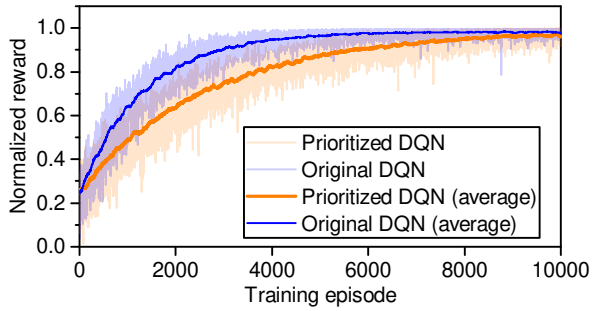

Fig. 14. Performance comparisons of training process.

DQN makes the training converge faster and has the slightly higher overall rewards.

In Section IV.B, the FRA-based DQN pruning method is proposed to reshape the DNN-based agent to make the agent more compact and reduce the computational complexity. The pruning process is only used for the fully connected layers. In this paper, before the training, the agent is designed as a neural network with two hidden layers. The first hidden layer includes 80 neurons and the second one includes 250 neurons. After training, the FRA-based pruning method is applied and the pruning process is performed based on Steps 1-4 in Section IV.B. The final RMSE of the loss is set to be lower than 0.2, which has been verified that the pruned agents will be guaranteed to perform the same actions as before. In the case studies, the DQN-based energy management is adopted in Scenarios 3, 4 and 5 respectively. Table VII shows the performance of the FRA-based pruning method for the trained agents in Scenarios 3, 4 and 5. The floating point operation (FLOPs) are considered to evaluate the computational cost. From Table VII, after pruning, the neuron numbers of both hidden layers of the three reshaped agents are significantly decreased compared to these of the original agents before pruning, with a maximum reduction of 73.6%. Accordingly, the FLOPs of both layers are also decreased and the maximum reduction is up to 87.2%. Compared to the original agents, the total parameters of the three reshaped agents are also decreased with reductions ranging from 51.7% to 79.2%. It is worth noting that the model sizes of the reshaped agents are also significantly reduced with the reductions of all three reshaped agents being more than 80.0% and a maximum reduction of 89.9%. The FRA-based agent pruning method not only makes the decision process faster, but also saves more memory resources for the controller. These benefits from the FRA-based pruning method are meaningful for the real application of the proposed DQN-based energy management.

VI. Conclusions

This paper has proposed a DQN-based thermal constrained energy management strategy for the railway co-phase system

TABLE VII
PERFORMANCE OF FRA PRUNING

		1st FC layer		2nd FC layer		Model size (KB)	Parameters
		Number of neurons	FLOPs	Number of neurons	FLOPs		
S3	Original	80	800	250	40,000	567	45,579
	Pruned	39	390	66	5,140	57	9,507
	Reduction	51.3%	51.3%	73.6%	87.2%	89.9%	79.1%
S4	Original	80	640	250	40,000	560	45,499
	Pruned	40	320	114	9,120	80	16,259
	Reduction	50.0%	50.0%	54.4%	77.2%	85.7%	64.3%
S5	Original	80	1,440	250	40,000	567	45,899
	Pruned	52	930	144	14,976	106	22,507
	Reduction	35.0%	35.4%	42.4%	62.6%	81.3%	51.0%

integrated with ESS. To constrain the thermal dynamics for both the power electronics modules and ESS and enhance their reliability, the overtemperature penalty terms are introduced in the problem formulation, and the DQN-based method is adopted to achieve the reactive compensation and peak traction load smoothing with the aid of the ESS. An FRA-based pruning method is then proposed to make the agent more compact without sacrificing the performance. Case studies are conducted to verify the effectiveness of the proposed method. The major conclusions are drawn as follows:

1) Without ESS integration, the proposed DQN-based energy management strategy for the thermal constraint outperforms the rule based strategies with the maximum thermal reductions of 89.3%. Meanwhile, the reactive power can be compensated as much as possible with the injected reactive power accounting for 9.0% of the maximum allowable amount.
2) With ESS integration, both DQN-based energy management strategies with or without considering thermal constraints can achieve the peak load shaving with maximum reductions of 31.9% and 42.2% respectively. Compared to the strategy without considering thermal constraints, the strategy considering thermal constraints can achieve a thermal reduction of 94.1% for power electronics modules and 100% for the ESS.
3) The FRA-based agent pruning method can achieve a maximum parameter reduction of 79.1% and a maximum agent size reduction of 89.9% without sacrificing the performance compared to original agents before pruning.

Finally, it needs to be pointed out that in practical applications, the operation of the railway co-phase power supply system is often subject to many uncertainties and stochasticity, and their impacts on the performance of the proposed DQN-based energy management strategy still deserve further studies. For example, forecasting techniques to capture the system stochasticity rooted in the uncertain traction demand and system parameter deviations need to be considered and incorporated into the proposed thermal-constraint energy management strategy.

REFERENCES

[1] I. A. Tasiu, Z. Liu, S. Wu, W. Yu, M. Al-Barashi, and J. O. Ojo, "Review of recent control strategies for the traction converters in high-speed train," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 2311–2333, 2022.

[2] L. Wang, Y. Pang, K.-W. Lao, M.-C. Wong, F. Ma, and X. Zhou, "Design and analysis of adaptive impedance structure for cophase railway traction supply power quality conditioner," *IEEE Transactions on Transportation Electrification*, vol. 6, no. 3, pp. 1338–1354, 2020.

[3] M. Chen, Z. Liang, Z. Cheng, J. Zhao, and Z. Tian, "Optimal scheduling of ftpss with pv and hess considering the online degradation of battery capacity," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 936–947, 2021.

[4] Z. Shu, S. Xie, and Q. Li, "Single-phase back-to-back converter for active power balancing, reactive power compensation, and harmonic filtering in traction power system," *IEEE Transactions on Power Electronics*, vol. 26, no. 2, pp. 334–343, 2010.

[5] S. Wu, M. Wu, L. Li, S. Wang, K. Song, and Y. Wang, "Analysis and comparison of mmc-based co-phase traction power supply topology for at power supply system," *IEEE Transactions on Power Delivery*, 2022.

[6] B. Xie, Y. Li, Z. Zhang, S. Hu, Z. Zhang, L. Luo, Y. Cao, F. Zhou, R. Luo, and L. Long, "A compensation system for cophase high-speed electric railways by reactive power generation of shc&sac," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 4, pp. 2956–2966, 2017.

[7] Y. Chen, M. Chen, Z. Liang, and L. Liu, "Dynamic voltage unbalance constrained economic dispatch for electrified railways integrated energy storage," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 8225–8235, 2022.

[8] M. Chen, Z. Cheng, Y. Liu, Y. Cheng, and Z. Tian, "Multitime-scale optimal dispatch of railway ftpss based on model predictive control," *IEEE Transactions on Transportation Electrification*, vol. 6, no. 2, pp. 808–820, 2020.

[9] İ. Şengör, H. C. Kılıçkıran, H. Akdemir, B. Kekezoğlu, O. Erdinç, and J. P. Catalão, "Energy management of a smart railway station considering regenerative braking and stochastic behaviour of ess and pv generation," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1041–1050, 2017.

[10] J. Wu, Z. Wei, W. Li, Y. Wang, Y. Li, and D. U. Sauer, "Battery thermal-and health-constrained energy management for hybrid electric bus based on soft actor-critic drl algorithm," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 3751–3761, 2020.

[11] M. Chen, M. Wang, D. Zhang, Y. Chen, and W. Lu, "Improved coordinated control strategy for reliability enhancement of parallel pfcs with lcc restriction," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 2093–2105, 2021.

[12] M. Chen, D. Zhang, M. Wang, Y. Lv, and Y. Chen, "A lifetime extension strategy to increase the reliability of pfc in co-phase tpss," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106969, 2021.

[13] R. Han, Q. Xu, H. Ding, P. Guo, J. Hu, Y. Chen, and A. Luo, "Thermal stress balancing oriented model predictive control of modular multilevel switching power amplifier," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9028–9038, 2019.

[14] A. Xuan, X. Shen, Q. Guo, and H. Sun, "A conditional value-at-risk based planning model for integrated energy system with energy storage and renewables," *Applied Energy*, vol. 294, p. 116971, 2021.

[15] A. Mao, T. Yu, Z. Ding, S. Fang, J. Guo, and Q. Sheng, "Optimal scheduling for seaport integrated energy system considering flexible berth allocation," *Applied Energy*, vol. 308, p. 118386, 2022.

[16] Y. Yerasimou, V. Pickert, B. Ji, and X. Song, "Liquid metal magnetohydrodynamic pump for junction temperature control of power modules," *IEEE Transactions on Power Electronics*, vol. 33, no. 12, pp. 10 583–10 593, 2018.

[17] K. Liu, K. Li, H. Ma, J. Zhang, and Q. Peng, "Multi-objective optimization of charging patterns for lithium-ion battery management," *Energy Conversion and Management*, vol. 159, pp. 151–162, 2018.
[18] J. Wu, Z. Wei, K. Liu, Z. Quan, and Y. Li, "Battery-involved energy management for hybrid electric bus based on expert-assistance deep deterministic policy gradient algorithm," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 12 786–12 796, 2020.
[19] J. Gonçalves, D. J. Rogers, and J. Liang, "Submodule temperature regulation and balancing in modular multilevel converters," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7085–7094, 2018.
[20] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
[21] Z. Wei, Z. Quan, J. Wu, Y. Li, J. Pou, and H. Zhong, "Deep deterministic policy gradient-drl enabled multiphysics-constrained fast charging of lithium-ion battery," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 3, pp. 2588–2598, 2021.
[22] S. Park, A. Pozzi, M. Whitmeyer, H. Perez, A. Kandel, G. Kim, Y. Choi, W. T. Joe, D. M. Raimondo, and S. Moura, "A deep reinforcement learning framework for fast charging of li-ion batteries," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 2, pp. 2770–2784, 2022.
[23] X. Tang, J. Zhang, D. Pi, X. Lin, L. M. Grzesiak, and X. Hu, "Battery health-aware and deep reinforcement learning-based energy management for naturalistic data-driven driving scenarios," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 948–964, 2021.
[24] B. E. Nyong-Bassey, D. Giaouris, C. Patsios, S. Papadopoulou, A. I. Papadopoulos, S. Walker, S. Voutetakis, P. Seferlis, and S. Gadoue, "Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty," *Energy*, vol. 193, p. 116622, 2020.
[25] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, and K. Li, "Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4513–4521, 2020.
[26] D. Livne and K. Cohen, "Pops: Policy pruning and shrinking for deep reinforcement learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 4, pp. 789–801, 2020.
[27] L. Wang, J. Xu, G. Wang, and Z. Zhang, "Lifetime estimation of igbt modules for mmc-hvdc application," *Microelectronics Reliability*, vol. 82, pp. 90–99, 2018.
[28] R. M. Schupbach and J. C. Balda, "Comparing dc-dc converters for power management in hybrid electric vehicles," in *IEEE International Electric Machines and Drives Conference, 2003. IEMDC'03.*, vol. 3. IEEE, 2003, pp. 1369–1374.
[29] C. Zhang, K. Li, and J. Deng, "Real-time estimation of battery internal temperature based on a simplified thermoelectric model," *Journal of Power Sources*, vol. 302, pp. 146–154, 2016.
[30] M. Seo, Y. Song, J. Kim, S. W. Paek, G.-H. Kim, and S. W. Kim, "Innovative lumped-battery model for state of charge estimation of lithium-ion batteries under various ambient temperatures," *Energy*, vol. 226, p. 120301, 2021.
[31] C. Zhang, K. Li, J. Deng, and S. Song, "Improved realtime state-of-charge estimation of lifepo$_4$ battery based on a novel thermoelectric model," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 1, pp. 654–663, 2016.
[32] C. Xing, K. Li, L. Zhang, and W. Li, "Optimal compensation control of railway co-phase traction power supply integrated with renewable energy based on nsga-ii," *IET Renewable Power Generation*, vol. 14, no. 18, pp. 3668–3678, 2020.
[33] D. J. Harrold, J. Cao, and Z. Fan, "Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning," *Energy*, vol. 238, p. 121958, 2022.
[34] Z. Yang, X. Ma, L. Xia, Q. Zhao, and X. Guan, "Reinforcement learning for fluctuation reduction of wind power with energy storage," *Results in Control and Optimization*, vol. 4, p. 100030, 2021.
[35] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[36] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv:1511.05952*, 2015.
[37] K. Li, J.-X. Peng, and G. W. Irwin, "A fast nonlinear model identification method," *IEEE Transactions on Automatic Control*, vol. 50, no. 8, pp. 1211–1216, 2005.
[38] K. Qu and J. Yuan, "Optimization research on hybrid energy storage system of high-speed railway," *IET Generation, Transmission & Distribution*, vol. 15, no. 20, pp. 2835–2846, 2021.
[39] *SKiiP IGBT Module 1213 GB123-2DL V3*, SEMIKRON, 3 2014.
[40] *SEMix IGBT Module 453 GB12-E4P*, SEMIKRON, 2 2020.
[41] X. Tang, J. Chen, H. Pu, T. Liu, and A. Khajepour, "Double deep reinforcement learning-based energy management for a parallel hybrid electric vehicle with engine start–stop strategy," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 1376–1388, 2021.
[42] Z. Shu, S. Xie, K. Lu, Y. Zhao, X. Nan, D. Qiu, F. Zhou, S. Gao, and Q. Li, "Digital detection, control, and distribution system for co-phase traction power supply application," *IEEE transactions on industrial electronics*, vol. 60, no. 5, pp. 1831–1839, 2012.
[43] N.-Y. Dai, M.-C. Wong, K.-W. Lao, and C.-K. Wong, "Modelling and control of a railway power conditioner in co-phase traction power system under partial compensation," *IET Power Electronics*, vol. 7, no. 5, pp. 1044–1054, 2014.