CHALMERS
UNIVERSITY OF TECHNOLOGY

Deep Deterministic Policy Gradient-DRL Enabled Multiphysics-Constrained Fast Charging of Lithium-Ion Battery

Citation for the original published paper (version of record):
Wei, Z., Quan, Z., Wu, J. et al (2022). Deep Deterministic Policy Gradient-DRL Enabled Multiphysics-Constrained Fast Charging of
Lithium-Ion Battery. IEEE Transactions on Industrial Electronics, 69(3): 2588 -2598.
http://dx.doi.org/10.1109/TIE.2021.3070514

Deep Deterministic Policy Gradient-DRL Enabled Multiphysics-Constrained Fast Charging of Lithium-Ion Battery

Zhongbao Wei, *Senior Member, IEEE*, Zhongyi Quan, *Member, IEEE*, Jingda Wu, Yang Li, *Member, IEEE*, Josep Pou, *Fellow Member, IEEE and* Hao Zhong

***Abstract*— Fast charging is an enabling technique for the large-scale penetration of electric vehicles. This paper proposes a knowledge-based, multi-physics-constrained fast charging strategy for lithium-ion battery (LIB), with a consciousness of the thermal safety and degradation. A universal algorithmic framework combining model-based state observer and a deep reinforcement learning (DRL)-based optimizer is proposed, for the first time, to provide a LIB fast charging solution. Within the DRL framework, a multi-objective optimization problem is formulated by penalizing the over-temperature and degradation. An improved environmental perceptive deep deterministic policy gradient (DDPG) algorithm with priority experience replay is exploited to trade-off smartly the charging rapidity and the compliance of physical constraints. The proposed DDPG-DRL strategy is compared experimentally with the rule-based strategies and the state-of-the-art model predictive controller to validate its superiority in terms of charging rapidity, enforcement of LIB thermal safety and life extension, as well as the computational tractability.**

***Keywords*— Fast charging, deep deterministic policy gradient, thermal safety, battery health, lithium-ion battery**

I. INTRODUCTION

Lithium-ion batteries (LIBs) have gained rapid popularity in electrified transportation due to their appealing features of high gravimetric and volumetric densities. Associated with the fast and foreseeable growth of electric vehicles (EVs) and LIB utilization, the past years have witnessed substantial research on battery management system (BMS), such as state estimation [1, 2], health prognostic [3, 4], and fault diagnostics [5, 6]. Charging of LIBs is recognized as a vital technology of future prosperity of EVs. However, the pursuit of utmost charging speed risks the violation of critical physical limits companied by the unexpected thermal/stress buildup and side reactions. Direct consequences of this include efficiency reduction, quick depletion, and even safety hazards in the most severe case.

Charging control has been a vast area of intensive studies, incubating a myriad of methods that can be categorized broadly into two groups. The first group is characterized with heuristic rule-based strategies which are model-free and widely adopted in real applications. Famous candidates include the constant-current-constant-voltage (CCCV) charging protocol [7] and a variety of variants, such as the multistage constant current (MCC) [8], multistage CCCV [9], and boost charging [10]. In spite of the low complexity, such methods are empirical without sufficient insight into the battery dynamics and physical constraints. Hence, such protocols are far away from optimality with respect to the charging speed and the enforcement of battery safety or longevity. This has motivated the exploration of the second group of methods, i.e., model-based strategies.

Model-based charging has the merit of more guaranteed optimality and higher robustness. The modeling techniques for LIB, which underlies this type of strategies, include the electrochemical model (EM) [11] and equivalent circuit model (ECM) [12]. Based on a coupled electro-thermal (CET) model, an optimized MCC strategy was proposed in [13], where the thermal and polarization effects were well confined. In [14], the CCCV strategy was optimized using a similar CET model and a multi-objective evolutionary approach. Within the same framework, the aging model was further incorporated to enable the health awareness [15]. Such methods plan the charging trajectory before the practical adoption via offline optimization, thus they are named as trajectory generator in this paper. It is feasible to embed the trajectory generator into the BMSs, where the user can select different charge patterns.

Unlike the aforementioned trajectory generator, model-based online controllers optimize the charging behavior in real time, and theoretically, they are most robust to the external disturbances. In particular, EMs were frequently used to describe the complex dynamics inside the LIB, enabling health-conscious fast charging control by either open-loop optimization [16], or online approaches such as proportional-

Manuscript received November 17, 2020; revised February 26, 2021, accepted March 21, 2021. This work was supported by the National Natural Science Foundation of China under Grant 52072038 (Corresponding author: Jingda Wu).

Z. Wei, J. Wu and H. Zhong are with the National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, China (e-mail: weizb@bit.edu.cn, jingda001@e.ntu.edu.sg, 3120205227@bit.edu.cn).

Z. Quan is with Department of Electrical and Computer Engineering, University of Alberta, Canada (e-mail: zquan@ualberta.ca).

Yang Li is with the School of Electrical Engineering, Chalmers University of Technology, Sweden (e-mail: yangli@ieee.org).

J. Pou is with the School of Electrical and Electronic Engineering, Nanyang Technology University, 639798 Singapore (e-mail: josep.pou@ieee.org).

integral-derivative (PID) control [17], nonlinear programming [18], or model predictive control (MPC) [19-21]. However, the intractable computation of nonlinear partial differential equations is a potential barrier for their real-world applications. To mitigate this challenge, a reduced-order EM was proposed to determine the limiting current of LIB in [22], which is insightful to the realization of fast charging control.

Compared to EMs, ECMs enjoy better computational tractability, thus have been used for charge optimization combing the objectives of fastness, limited temperature buildup, and health retention [23, 24]. Within similar frameworks, the user-involved optimal charging was further achieved in [25] by enabling the objective specification. Recently, a hierarchical architecture combining ECM-based offline trajectory generator with online path tracking controller has been proposed, which allows a cost-effective charge control of both battery cells and packs [26, 27]. Most recently, an ECM-based explicit MPC controller was proposed for LIB fast charging, to reduce the complexity rooted in the constrained optimization [28].

The model-based charging strategies have two major drawbacks. First, they are sensitive to the accuracy of the battery model, while an intrinsic paradox is that an improved accuracy always compromises the computational tractability. Second, even by using reduced-order models accounting only for the lumped dynamics, the computation is still expensive due to the need of nonlinear optimization. A fast charging approach with the merits of both multi-objective optimality and online tractability is thereby highly desired.

Reinforcement learning (RL) is an efficient machine learning approach used for solving a broad range of optimization problems. Unlike the supervised/unsupervised learning, RL algorithms give memorable feedback on the cost function and search for the optimal solution automatically [29]. Attributed to the end-to-end characteristic, a high potential can be expected for the RL to be used on optimal charging control. The model-free feature is also favorable for ruling out the model sensitivity problem of model-based methods. RL-related studies have been disclosed in the field of charging plan of plug-in electric cars [30], vehicle charging station management [31], and the energy storage arbitrage [32]. Nonetheless, RL-based EV charging optimization is still in infancy. The exploitation of RL in LIB fast charging with thermal and aging consciousness has never been attempted beforehand.

This paper bridges the aforementioned gaps and proposes a novel knowledge-based, multi-physics-constrained fast charge strategy for LIBs. The strategy consists of an observer for state of charge (SOC) and internal temperature joint estimation, and a deep RL (DRL) controller for thermal- and health-aware fast charging. Four primary contributions are made.

First, the DRL is introduced for the first time to solve the LIB fast charging problem. A universal algorithmic framework incorporating the model-based state observer and the learning-based optimizer is proposed.

Second, a multi-constrained least costly objective is formulated by augmenting penalties for the over-temperature and degradation, to allow accounting for the thermal safety and life fading of LIB during the charging control.

Third, an environmental perceptive, fast-converging deep deterministic policy gradient (DDPG) algorithm, with priority experience replay, is exploited to improve the performance of multi-objective optimization in the formulated framework.

Lastly, unlike most of fast charging works that use real-time simulation for validation, real-world long-term experiments are performed to validate the proposed strategy more faithfully.

The contributions eventually give rise to a smart, thermal- and health-aware fast charging strategy. To the best of our knowledge, this is the first attempt to use machine learning techniques for the fast charging of LIB.

The remainder of the paper is organized as follows. An electro-thermal-aging model of LIB is presented in Section II. Section III details the proposed DDPG-DRL strategy. Results are discussed in Section 0, while the major conclusions are drawn in Section V.

II. Battery Modeling

A. Electro-Thermal Modeling for LIB

A coupled electro-thermal model is established, as shown in Fig. 1, to predict the electrical and thermal dynamics of the investigated A123 $LiFePO_4$ cylindrical battery. The model comprises a second-order RC model and a two-state thermal model. In terms of the electrical model, the voltage source describes the SoC-dependent open-circuit voltage, while R_s is the ohmic resistance. The two RC branches simulate polarization effects including charge transfer, diffusion, and passivation layer effect on electrodes. The governing equations of the second-order RC model are given by:

$$\frac{dSoC(t)}{dt} = \frac{I(t)}{3600C_n} \tag{1}$$

$$\frac{dV_{p1}(t)}{dt} = -\frac{V_{p1}(t)}{R_{p1}(t)C_{p1}(t)} + \frac{I(t)}{C_{p1}(t)} \tag{2}$$

$$\frac{dV_{p2}(t)}{dt} = -\frac{V_{p2}(t)}{R_{p2}(t)C_{p2}(t)} + \frac{I(t)}{C_{p2}(t)} \tag{3}$$

$$V_t(t) = V_{oc}(SoC,t) + V_{p1}(t) + V_{p2}(t) + R_s(t)I(t) \tag{4}$$

where I is the load current, V_t is the terminal voltage, C_n is the nominal capacity of the battery, and V_{p1} and V_{p2} are the polarization voltage across the two RC branches.

The thermal-energy conservation principle defines:

$$\frac{dT_s(t)}{dt} = -\frac{T_s(t)}{R_uC_s} - \frac{2T_s(t)}{R_cC_s} + \frac{2T_a(t)}{R_cC_s} + \frac{T_f}{R_uC_s} \tag{5}$$

$$\frac{dT_a(t)}{dt} = \left(\frac{C_s - C_c}{R_cC_cC_s} - \frac{1}{2R_uC_s}\right)T_s(t) + \frac{C_c - C_s}{R_cC_cC_s}T_a(t) + \frac{H(t)}{2C_c} + \frac{T_f}{2R_uC_s} \tag{6}$$

where T_s, T_a and T_f are battery surface, internal average, and ambient temperature, respectively, R_c and R_u are thermal resistances due to the heat conduction inside the battery and the convection at battery surface, C_c and C_s are equivalent

thermal capacitances of the battery core and surface. H is the heat generation rate.

Specifically, the heat is generated from three sources, i.e., the ohmic heat, polarization heat and the irreversible entropic heat. The heat generation rate can be calculated by:

$$H(t) = I(t)[V_{p1}(t) + V_{p2}(t) + R_s(t)I(t)] + I(t)[T_a(t) + 273]E_n(SoC,t) \quad (7)$$

where E_n denotes the entropy change during electrochemical reactions. Subsequently, the core temperature is given by:

$$T_c(t) = 2T_a(t) - T_s(t) \quad (8)$$

The employed model is widely explored in the literature, and thus the values of involved model parameters are not elaborated herein for briefness. However, more details can be referred to [33], where the calibration environment, protocols, and results are given systematically.

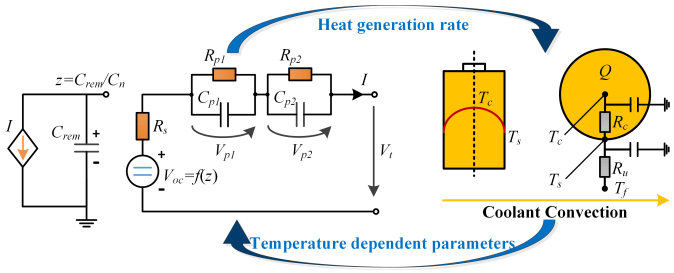


Fig. 1. Electro-thermal model of A123 $LiFePO_4$ cylindrical battery.

B. Aging Model of LIB

The energy-throughput-based model has been well validated for the A123 $LiFePO_4$ (26650) cylindrical battery in use [34], thus is used to quantify the capacity loss herein. The throughput model assumes the LIB can withstand a certain amount of charge flow, equivalent to cycles of charge and discharge, before it reaches the end-of-life.

The C-rate (c) and battery internal temperature have large impact on the capacity fade. The Arrhenius equation-based capacity loss is given by:

$$\Delta C_n = B(c) \cdot e^{\frac{-E_a(c)}{RT_a}} \cdot Ah(c)^z \quad (9)$$

where ΔC_n is the percentage of capacity loss, B the C-rate-dependent pre-exponential factor which can be referred to TABLE I [34], R the ideal gas constant, z the power-law factor equals to 0.55, Ah the accumulated ampere-hour throughput, and E_a the activation energy (J / mol) defined by:

$$E_a(c) = (31700 - 370.3 \cdot c) \quad (10)$$

LIBs reach the end-of-life when C_n drops by 20%. Being aware of this, and referring to (9), Ah and the total cycling number before reaching the end-of-life (N) can be derived as:

$$Ah(c,T_a) = \left[20 \Big/ B(c) \cdot \exp\left(\frac{-E_a(c)}{RT_a} \right) \right]^{1/z} \quad (11)$$

$$N(c,T_a) = 3600 Ah(c,T_a) / C_n \quad (12)$$

Afterward, the drop of state of health (SoH) under multiple stresses is given by:

$$\Delta SoH_k = -\frac{|I_k| \Delta t}{2N_k(c,T_a)C_n} \quad (13)$$

where Δt is the lasting time of current.

TABLE I
DEPENDENCE OF PRE-EXPONENTIAL FACTOR TO C-RATE

c	0.5	2	6	10
$B(c)$	31630	21681	12934	15512

III. FAST CHARGING STRATEGY

The proposed charging strategy is comprised of a model-based observer used for estimating the unmeasurable states of LIB, and a DDPG-DRL optimizer for online charging control. The involved sub-algorithms are elaborated in this section.

A. State Observer

A model-based state observer is devised to estimate the unmeasurable SoC and T_c and thus enable the state-feedback control framework. The electro-thermal functions (1)-(8) are utilized to build a state-space model, where the state variables are V_{p1}, V_{p2}, SoC, T_c, and T_s, the system input is I, while the system outputs are V_t and T_s. Considering the nonlinearity of the system, an extended Kalman filter (EKF) is used to design the state observer in this paper. The algorithmic procedures of EKF are summarized in TABLE II [35], where Σ_w and Σ_y are the covariance matrix of process and measurement noises.

TABLE II
ALGORITHMIC PROCEDURE OF EKF

Initialization: $\hat{x}_0$, P_0, Σ_w, Σ_v
Definition: $F_k = \frac{\partial f(x(k-1),u(k-1))}{\partial x}$, $G_k = \frac{\partial h(x(k),u(k))}{\partial x}$
for $k = 1, 2, \ldots$
Prior state update: $\hat{x}(k) = f\left(x(k-1),u(k-1)\right)$
Prior error covariance update: $P_k = F_k P_{k-1} F_k^T + \Sigma_w$
Kalman gain update: $K_k = P_k G_k^T (G_k P_k G_k^T + \Sigma_v)^{-1}$
Posteriori state update: $x(k) = \hat{x}(k) + K_k\left(z(k) - g(\hat{x}(k),u(k))\right)$
Posteriori error covariance update: $P_k = (I - K_k G_k) P_k$

B. Optimization Problem Formulation

It is favorable that a charging solution can make an optimal balance among conflicting objectives of charging rapidity, thermal safety, and life extension. In this paper, the optimal control is realized by minimizing the following cost function:

$$J_t = \omega_1 C_{soc} + \omega_2 C_{volt} + \omega_3 C_{heat} + \omega_4 C_{soh} + \omega_5 C_{smooth} \quad (14)$$

where ω_1, ω_2, ω_3, ω_4 and ω_5 are weights describing the importance of different targets.

C_{soc} describes the charging time and is given by:

$$C_{soc} = \left| SoC_{tar} - SoC_t \right| \quad (15)$$

where SoC_t and SoC_{tar} denote the present SoC and the target SoC at the end of charge. The RL agent sets the expectation on overall rewards with respect to future time steps as its objective. Hence, this term means that an action suppressing this deviation (high current) will be awarded, while a conservative action causing large deviation (low current) will be penalized. In this way, the RL agent is guided to pursuit high charging currents during the training, and thus shortens the charging time.

C_{volt} and C_{heat} denote the safety-violating cost with respect to over-voltage and over-temperature of LIB. Particularly, the terminal voltage and internal temperature of LIB are desired to be controlled below a specific threshold. Using hard constraints potentially disrupt the exploration process of DRL considering the high possibility of constraint violation. Therefore, the following soft penalties are instead employed:

$$C_{volt}=\begin{cases}0 & if \quad V_{tar\_low}<V_t<V_{tar\_upp}\\ \tau_1\left|V_t-V_{tar\_upp}\right| & if \quad V_t>V_{tar\_upp}\\ \tau_1\left|V_t-V_{tar\_low}\right| & if \quad V_t<V_{tar\_low}\end{cases} \tag{16}$$

$$C_{heat}=\begin{cases}0 & if \quad T_{a,t}<T_{tar}\\ \tau_2\left|T_{a,t}-T_{tar}\right| & if \quad T_{a,t}\geq T_{tar}\end{cases} \tag{17}$$

where V_t, V_{tar_upp}, and V_{tar_low} are the present, upper and lower limit terminal voltage, respectively, $T_{a,t}$ and T_{tar} are the present and upper limit internal temperature of LIB, respectively. The over-voltage should be avoided to ensure the safety of LIB in practical applications. Hence, a large weighting factor (three order of magnitude higher than the rest) is used for ω_2 to ensure a compliance to the voltage constraint without overshooting.

C_{soh} denotes the aging cost of LIB given by:

$$C_{soh}=\tau_3\left|\Delta SoH_t\right| \tag{18}$$

where ΔSoH_t is the drop of SOH as a consequence of the present control action. Note that τ_1, τ_2 and τ_3 are transition coefficients enforcing C_{volt}, C_{heat} and C_{soh} dimensionally comparable to C_{soc}.

The charging current is desired to be controlled smoothly in real applications. In this regard, C_{smooth} describing the cost of control effort is given by:

$$C_{smooth}=\left|I_t-I_{t-1}\right| \tag{19}$$

C. Improved DDPG Algorithm

Derived from the actor-critic structure, the DDPG algorithm is devised with two deep neural networks (DNNs), i.e., a value (critic) network Q and a policy (actor) network μ. The policy-network behaves as an actor to map the state-space composition to a continuous action δ, while the value-network behaves as a critic, which timely evaluates the policy function's performance and gives feedback for improvement. Target networks Q' and μ' are used to track the original Q and μ network, so as to mitigate the effect of incorrect evaluation. Note that the target networks possess the same structures and initial weights, yet update the network parameters more robustly.

The determination of action of DDPG in a specific timestep t considered both exploration and the inherent policy, which is given by:

$$a_t=\mu(s_t\,|\,\theta^{\mu})+\varepsilon \quad \varepsilon\sim\mathcal{N} \tag{20}$$

where s_t is the state space, and θ^{μ} is the parameters of network μ, and ε is the Gaussian noise which exists only in the training stage.

Since the policy is determined in the training stage, the principles of offline training are clarified hereafter. The policy evaluation is performed based on the Bellman's principle as:

$$Q^*(s_t,a_t)=E[r(s_t,a_t)+\gamma\arg\max_{a_t}(Q^*(s_{t+1},a_{t+1}))] \tag{21}$$

where Q^* denotes the optimal value function, r the single-step reward and γ the discount factor.

Equation (21) reveals that the optimal evaluation of present composition of states and actions can be obtained recursively. It is expected that the deep networks Q and Q' can approximate this iterative task accurately. To realize it, the updating error of value network Q can be calculated by:

$$L_Q(t\left|\theta^Q\right.)=\left[\left(r(s_t,a_t)+\gamma Q'(s_{t+1},a_{t+1}\left|\theta^{Q'}\right.)-Q(s_t,a_t\left|\theta^Q\right.)\right)\right]^2 \tag{22}$$

$$a_{t+1}=\mu'(s_t\left|\theta^{\mu'}\right.) \tag{23}$$

where the first two terms in (22) denote the expected Q value referring to (21), and the last term refers to the actual output of current value network. In this way, the squared error can be obtained, and the gradient-descent updating method can be performed to improve the policy evaluation ability.

An ideal value network is expected to output the accurate evaluation of policy, so that the actor network can adjust its policies accordingly to discard the actions with bad Q value feedback. Therefore, the performance objective of policy network, represented by ϕ, can be defined as:

$$\phi(\theta_\mu)=E[-Q(s_t,\mu(s_t))] \tag{24}$$

where $E(\cdot)$ denotes the expectation operator.

The policy network keeps updating itself towards the direction of promoting the performance objective. Therefore, the updating error can be expressed as the gradient of objective with respect to network μ:

$$L_\mu(t\left|\theta^\mu\right.)=\nabla_{\theta^\mu}\phi(\theta^\mu)=\nabla_a Q(s_t,\mu(s_t)\left|\theta^Q\right.)\nabla_{\theta^\mu}\mu(s_t\left|\theta^\mu\right.) \tag{25}$$

A soft updating strategy is adopted for the target networks Q' and μ', given by:

$$\begin{aligned}\theta^{Q'}&\leftarrow\tau\theta^Q+(1-\tau)\theta^{Q'}\\ \theta^{\mu'}&\leftarrow\tau\theta^\mu+(1-\tau)\theta^{\mu'}\end{aligned} \tag{26}$$

The experience replay method is further adopted for the DDPG algorithm to avoid the back-forth correlation of trained networks. Different from the simple random sampling adopted by conventional DDPG, the improved DDPG algorithm endows the importance weights to experience sample. This mechanism is inspired from the fact that the highly rewarded or painful experiences are more informative than the plain ones. The experience replay method, which emphasizes those impressive experiences, is hence expected to improve the efficiency and stability of learning.

The probability of the sampled experience j can be described as:

$$\begin{aligned}P_j&=D_j^\alpha\Big/\left(\Sigma_k D_k^\alpha\right)\\ D_j&=1/rank(j)\end{aligned} \tag{27}$$

where $\sum_k(\cdot)$ denotes the total index in the experience pool, and α is the hyperparameter to determine priority degree, ranging from 0 to 1. Lower α tends to uniform sampling of conventional DDPG, $rank\ (\cdot)$ is the importance degree of a set of experience, which can be calculated by:

$$rank(j)=\sqrt{L_Q(j)} \tag{28}$$

By adopting the experience replay, those experiences causing more significant changes to the policy evaluation will be assigned more weights, and therefore, are more likely to be chosen and replay in the training process.

D. Continuous DRL-Based Charging Control

To solve the optimization problem suggested by (14) in the continuous DRL framework, the reward function has to be expressed alternatively by:

$$r(s_t, a_t) = f_{nor}(b - J_i) \tag{29}$$

where b is a user-defined bias to adjust the range of reward function, and $f_{nor}(\cdot)$ denotes a sigmoid-based normalization function, which contributes to consolidating the physical variables into a unified range of [-1,1].

In this work, the state space is defined as:

$$S = \left\{SoC, f_{nor}(T_c), f_{nor}(V_t)\right\} \tag{30}$$

where V_t is directly measurable, while SoC and T_c can be estimated online using model-based observer in Section III-A.

The DDPG-DRL strategy is expected to control the charging current in a continuous manner, thus the action space can be defined as:

$$A = \{c_t \,|\, c_t \in (0, 6C)\} \tag{31}$$

where the upper limitation of 6 C is determined based on the specification of the investigated LIB.

With the afore-defined reward function, state and action space, the architecture of the DDPG-DRL fast charging strategy has been put forward. Particularly, the diagram of the DDPG-DRL strategy is shown schematically in Fig. 2, while the associated hyperparameters are listed in Table IV. For clarity, the procedures of training and real-time implementation are detailed in TABLE IV and TABLE V, respectively.

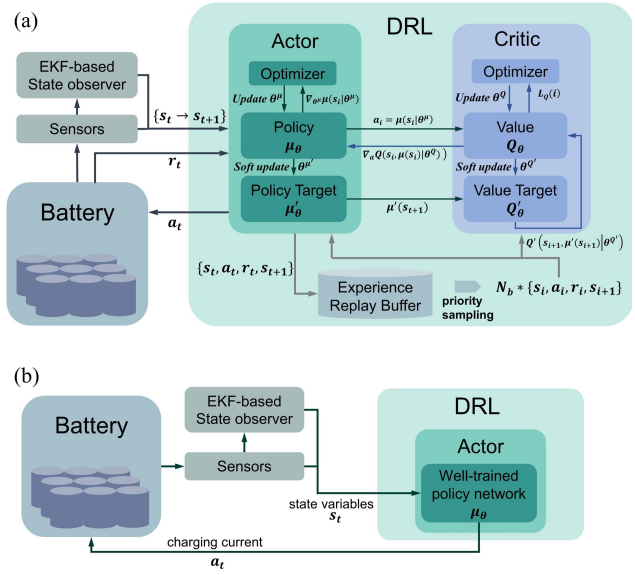


Fig. 2. Implementation of the DDPG-DRL fast charging strategy. (a) training process and its principles, and (b) real-time application process.

TABLE III
HYPERPARAMETERS USED FOR THE PROPOSED DDPG-DRL STRATEGY

Parameter	Description (unit)	Value
N_E	Experience pool size	384000
M	Training total steps	500000
T_l	Maximum episode length (s)	2000
N_b	Minibatch size	128
lr^a	Initial learning rate (policy network)	0.001
lr^c	Initial learning rate (value network)	0.002
γ	Discount factor	0.99
τ	Soft updating factor	0.01

TABLE IV
TRAINING PROCEDURE OF DDPG FAST CHARGING STRATEGY

DDPG-based ageing- and heating- aware fast charging algorithm

1. Inputs: initial policy parameters θ^μ and $\theta^{\mu'}$, value parameters θ^Q and $\theta^{Q'}$, empty experience replay buffer D.
2. *While* $epoch < threshold$:
   Initialize the battery model.
   While not terminate:
   Obtain the state space S_t, which is consisted of normalized state variables s_t, from the battery model.
   Select action $a = \mu(s_t) + \varepsilon$, mapping the action into the expected charging current I_t.
   Execute the I_t in the battery model.
   Observe next state s_{t+1}, reward r_t
   Store transition $\{s_t, a_t, r_t, a_{t+1}\}$ in the priority experience buffer D.
   Retrieve a batch of transitions, $B = \{s_i, a_i, r_i, s_{i+1}\}$ from D according to the probability of the priority experience mechanism.
   Update the value network with:
   $L_Q(t|\theta^Q) = \left[\left(r(s_t, a_t) + \gamma Q'(s_{t+1}, a_{t+1}|\theta^{Q'}) - Q(s_t, a_t|\theta^Q)\right)\right]^2$
   Update the policy network with:
   $L_\mu(t|\theta^\mu) = \nabla_{\theta^\mu}\phi(\theta^\mu) = \nabla_a Q(s_t, \mu(s_t)|\theta^Q)\nabla_{\theta^\mu}\mu(s_t|\theta^\mu)$
   Update the target networks with:
   $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
   $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
   If s_{t+1} triggers the episode terminated condition:
   $epoch = epoch + 1$
3. Save parameters of the policy network θ^μ for real-time applications.

TABLE V
REAL-TIME CONTROL PROCEDURE OF DDPG FAST CHARGING STRATEGY

DDPG-based ageing- and heating- aware fast charging algorithm

1. Construct the state observer, config the input/output of the LIB system.
2. Load the trained parameters of the policy network of the DDPG agent, set constrained thresholds for the input/output variables of the policy network.
3. *While* not terminate:
   Send the state variable of the present time step s_t to the policy network
   Obtain the network's output of the present time step a_t.
   Map the network's output into the expected charging current I_t.
   Check if the termination condition is satisfied.

IV. RESULTS AND DISCUSSION

A. Validation of Battery Modeling

The A123 26650 LIB cell is cycled with 2 C, 4 C and 6 C using Arbin testing system, which consists of the programmable electrical load and power supply. The ranges of sensors are 10 A and 5 V, while the error limits are both within 0.05%. The test cell is placed in a programmable thermal chamber to control the ambient temperature at 25°C during the experiment. At the same time, three thermocouples are attached to different surface locations of the cylindrical cell along the axial direction, and the averaged readings of them are treated as the surface temperature. The modeled battery terminal voltage and surface temperature are plotted against their experimental benchmarks in Fig. 3. It is shown that the modeled results resemble the ground truth closely at different C-rates. The corresponding statistical errors are summarized in TABLE VI. The observed

low modeling errors validate the high fidelity of the presented model for describing the electro-thermal dynamics of LIB.

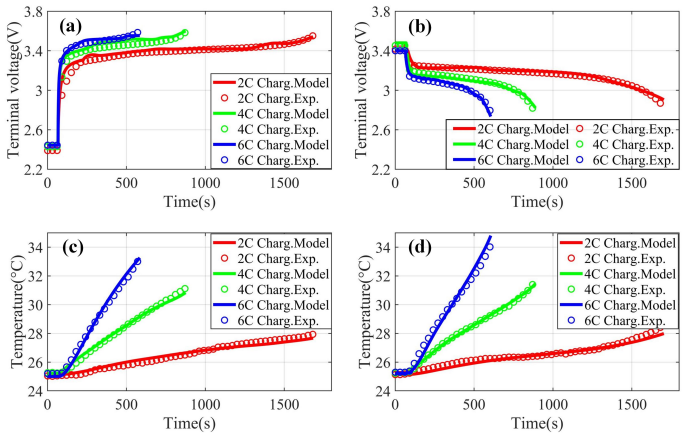


Fig. 3. Results of model validation: terminal voltage of (a) charge, and (b) discharge, surface temperature of (c) charge, and (d) discharge.

TABLE VI
MODELING ERRORS AT DIFFERENT C-RATES

	Terminal voltage (V)			Surface temperature (°C)		
	2 C	4 C	6 C	2 C	4 C	6 C
MAE	0.0199	0.0294	0.0146	0.150	0.080	0.260
RMSE	0.0297	0.0349	0.0242	0.164	0.102	0.296

B. Validation of Training Process

The training performance as a key measure of the proposed DDPG-DRL charging strategy is evaluated in this section. The episodic average reward value is illustrated in Fig. 4 (a). Explicitly, an increased reward value implies the improvement of trained charging strategy from the optimality point of view. The physical indicators are depicted in Fig. 4 (b-d) for further validation. It is shown that the mentioned early termination is attributed to the overcharging, i.e., the end SOC exceeds the upper threshold. As the training proceeds, the end SOC is suppressed towards the target SOC, which reveals the compliance to the constraints. Meanwhile, the terminal voltage and battery temperature are both confined to reasonable levels. All the results have validated the convergence and potential feasibility of the trained policy.

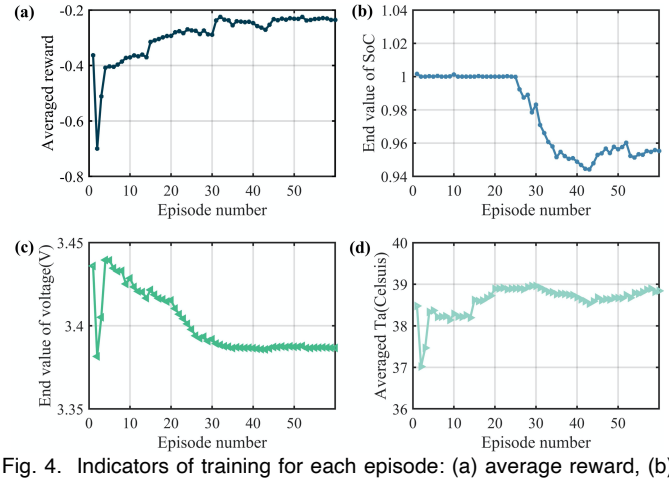


Fig. 4. Indicators of training for each episode: (a) average reward, (b) end SOC, (c) average terminal voltage, and (d) average cell temperature.

C. Thermal and Health-Conscious Validation: Simulation

The proposed strategy manifests itself with the LIB over-heat protection and life extension by penalizing the high temperature and degradation in the cost function. To justify this merit, it is compared with a baseline strategy, i.e., its counterpart without thermal and health constraints, while the other configurations are kept consistent. To rule out the effect of model uncertainty and give a theoretical validation, the strategies are carried out in a simulation environment herein. In particular, the presented electro-thermal-aging model is used as a "virtual battery" and implant to the OPAL-RT real-time simulator, while the strategies are executed with the embedded processor. It is shown that the occupied execution cycle is only 5.45 μs, and no overrun is reported, which validates the real-time tractability.

The comparative results are shown in Fig. 5. It is shown that the proposed strategy needs 699 s to charge the LIB to the target SOC, which is 4.43% longer than using the baseline strategy. The more conservative charging is rooted in the restriction of charging current to ensure the expected thermal and degradation performance. As shown in Fig. 5 (c), the internal temperature of LIB increases to over 45°C by using the baselined strategy. In contrast, the proposed strategy keeps the internal temperature well below the defined threshold. This is within expectation as a high temperature introduces extra "cost" by the penalty imposed, while an excessively low temperature compromises the charging speed inevitably.

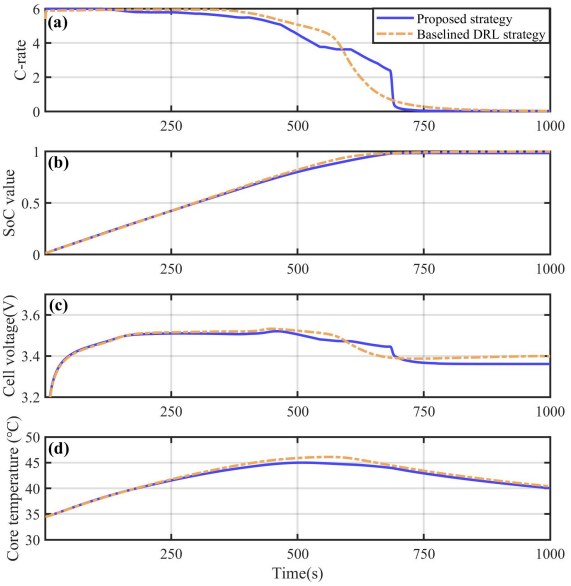


Fig. 5. Comparative results of DDPG-DRL strategies: (a) SOC, (b) terminal voltage, (c) core temperature, and (d) charging current of LIB.

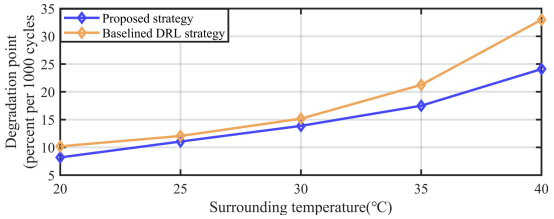


Fig. 6. The SOH drops by using different DDPG-DRL strategies.

A long-term simulation consisting of 1000 charging cycles is performed to evaluate the proposed strategy in terms of battery life extension. The SOH drops by using different strategies under different surrounding temperatures are illustrated in Fig. 6. Three conclusions can be drawn. First, as expected, the degradation accelerates with elevated temperature, due to the enhanced aging modes like the SEI growth. Second, the proposed strategy suppresses the aging rate compared to the baseline strategy, attributed to the penalties to over-temperature and quick degradation. Third, the anti-degradation potential of the proposed strategy becomes more prominent with the rise of surrounding temperature. It can be inferred that within a colder environment, the temperature rise is not sufficient to trigger the over-temperature penalty, so that the difference is hardly observable. In contrast, the temperature easily breaks the upper limitation under a relatively high temperature like 40°C. In this case, both the two penalizing mechanisms in the proposed strategy take effect, leading to a superimposed and thus stronger effect of anti-degradation.

D. Comparison of Strategies: Experimental Validation

This section goes further to compare the proposed strategy with the state-of-the-art benchmarks, i.e., the rule-based and model-based ones. It is worth noting that the model mismatch can decline the performance of the proposed strategy in practice. The strategies are hence applied on real-world batteries for experimental validation. The validation environments are consistent for different strategies to ensure a fair comparison.

The CCCV charging as a rule-based strategy is most-widely used in practical applications. The experimental charging results by using 2C, 4C, 6C CCCV strategies and the proposed strategy are shown comparatively in Fig. 7. The time consumed to charge the LIB to different charge levels, i.e., 80%, 90% SOC and fully charge, are summarized in TABLE VII. With respect to the CCCV strategies, it is explicit that a trade-off always exists between the charging speed and the threshold compliance. Although the 2C CCCV strategy ensures a favorable thermal condition, its charging is much slower than the other strategies. As the CC rate increases to 6 C, the charging time has been reduced largely. However, the accelerated charging is at the expense of over-temperature, which is unfavorable from the safety and longevity point of view. It is hence concluded that the CCCV strategy is far from optimality, since it fails to control the charging smartly to fulfill multiple objectives.

By comparison, the DDPG-DRL strategy shows to manage the trade-off smartly. It is shown that the estimated internal temperature of LIB is well confined to the threshold of 45°C, which is quite similar to the case of 4C CCCV strategy. However, its charging time is 36.7%, 33.7% and 28.6% shorter than the 4C CCCV strategy for the three end charging points. Compared to the 6C CCCV strategy, the charging time is quite approaching, but the risk of battery over-heat is strictly avoided. Overall speaking, the proposed DDPG-DRL strategy succeeds to find a balanced solution between the 4C and 6C CCCV strategy by accounting for conflicting objectives of both the charging rapidity and the physical constraint compliance.

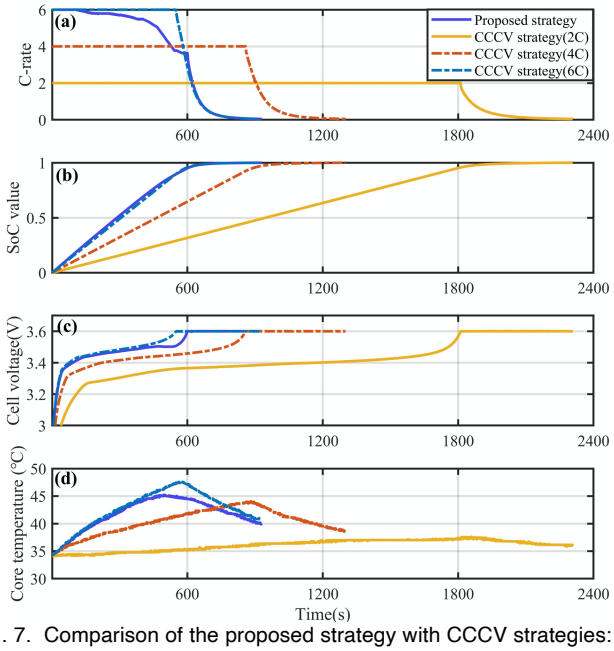


Fig. 7. Comparison of the proposed strategy with CCCV strategies: (a) current, (b) SOC, (c) terminal voltage, and (d) LIB core temperature.

TABLE VII
COMPARISON OF CHARGING SPEEDS FOR DIFFERENT STRATEGIES

Strategy	DDPG-DRL	2C CCCV	4C CCCV	6C CCCV	MPC
To 80% SOC, s	475	1515	743	490	521
To 90% SOC, s	554	1703	836	552	608
Fully charge, s	926	2283	1297	914	1054

The MPC as a typical model-based optimization method is further compared with the proposed DDPG-DRL strategy, and the experimental results are shown in Fig. 8. It is shown that the charging currents given by the two strategies follow a similar trajectory, i.e., maintaining at the highest allowable value at early stage while heading downwards as the charging proceeds, to keep the critical variables within expected ranges. As seen from TABLE VII, the charging speed is similar for the two strategies. Moreover, the LIB internal temperature is controlled well within the imposed thresholds for both of the two strategies. It is worth noting that the MPC gives a more conservative solution, witnessed by the under-shot temperature against the threshold of 45°C and the slightly longer charging time. This is more likely caused by the model mismatch which distorts the control trajectory to some extent compared to the ideal condition. Such slight deviations, however, cannot promise any virtual difference of the two strategies. It is thereby validated that the DDPG-DRL strategy performs equivalently with the state-of-the-art MPC strategy. Despite the similar optimality, the online tractability of DDPG-DRL strategy is much more favorable than the MPC, which will be discussed in detail in following sections.

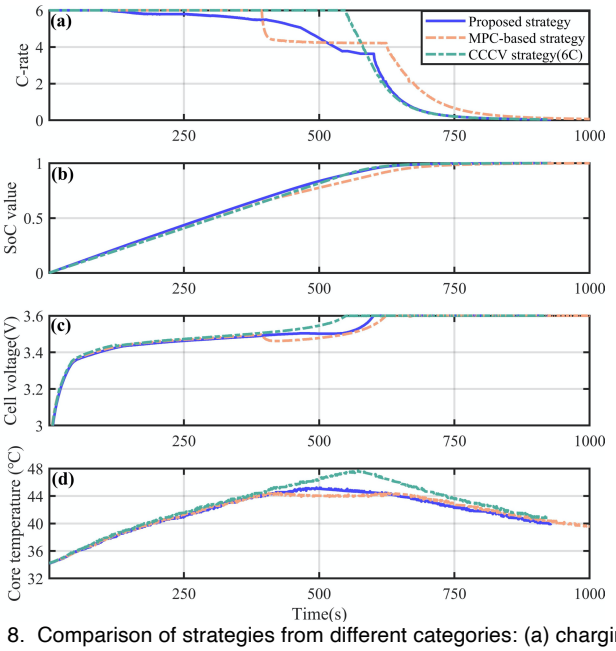


Fig. 8. Comparison of strategies from different categories: (a) charging current, (b) SOC, (c) terminal voltage, and (d) LIB core temperature.

Long-term cycling experiments are further performed to evaluate the health-conscious properties of different strategies. The candidates for comparison herein include the 6C CCCV, MPC, and DDPG-DRL strategy. The selection is made based on the fact that, these strategies share similar charging speeds, thus the difference in health degradation rate can be a strong measure of their optimality. Specifically, the strategies are applied on real-world batteries for charging, while a consistent 0.3 C discharge is applied to deplete the LIB. The described cycles are repeated to observe the causal effects of different strategies on the battery degradation.

The results of LIB capacity fade using different strategies are summarized in TABLE VIII. Explicitly, the experimental and calculated SOH drops disclose a consistent trend over different strategies, albeit an average deviation of 0.30% exists between the modelling and experiment due to the intrinsic error of aging model. The observed error is acceptable as the highly nonlinear aging path of LIB is extremely difficult for accurate modeling. From the health-conscious point of view, the proposed strategy and MPC show lower degradation rates, attributed to the well-constrained temperature and C-rate. By comparison, the 6C CCCV strategy incurs vastly faster degradation. Specifically, the proposed strategy can elongate the LIB life time by 14.8% compared to the 6C CCCV strategy when the charging speed is approximately equivalent. The faster LIB degradation under 6C CCCV mode can be explained by that the aging-dependent stress variables have been left unregulated

To attest this conjecture, the operating points distributed in the aging severity factor map are plotted comparatively in Fig. 9 (a). It is shown that the operating points are more likely distributed at the upper right quarter of the map with high aging severity by using 6C CCCV strategy. By observing the boxplot of severe factor in Fig. 9 (b), the 6C CCCV strategy gives rise to an average severe factor of 5.46, while the highest severe factor reaches up to 7.88. By comparison, the DDPG-DRL and MPC strategy control the average severe factor at 5.15 and 5.13, respectively, suggesting a much-relieved aging stress. These results reveal the distinct aging paths of LIB, which is the underlying reason of life extension of the proposed strategy. In summary, the slower aging with similar charging speed well supports the superiority of DDPG-DRL and MPC strategy.

Since the training of strategy is based on the built electro-thermal-aging model, any model mismatch can be transferred to the optimality of strategy. Therefore, meticulous model parameter calibration should be performed before the training to provide a mathematical guarantee on the control performance. With respect to the present case, moderate model deviations can be observed in Fig. 3 and TABLE VI due to the errors of parameters. In accordance, evident differences exist between experimental results (Fig. 7, Fig. 8 and TABLE VIII) and simulation results (Fig. 5 and Fig. 6). However, the practical control validates to guarantee an expected performance in the charging rapidity, as well as the thermal and health protection. In the severest case, slight constraint violation can occur due to the model mismatch, but this can be easily corrected by pre-set rules to comply better to the constraints.

TABLE VIII
SOH DROPS FOR 100 CHARGING CYCLES USING DIFFERENT STRATEGIES

SOH drop	DDPG-DRL	MPC	6C CCCV
Experimental	0.88%	0.86%	1.01%
Calculated	1.21%	1.17%	1.27%

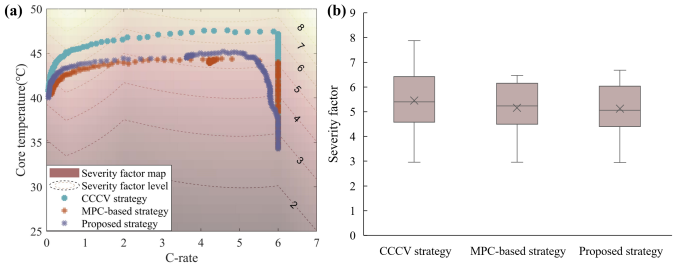


Fig. 9. Comparison of anti-aging performance: (a) operating points distributed in severity factor map, and (b) boxplot of severity factors.

E. Computing Complexity

The computational complexity is critical to evaluate the feasibility of strategies in practical applications. The counting of floating-point operations is known as a crude method to measure the order of computational complexity via the big-O-notion. In this regard, the MPC controller has cubic complexity considering the need of matrix multiplication and inversion. The multi-step optimization task within the control horizon further aggregates the numerical complexity. By comparison, the vast majority of computing cost of DDPG-DRL strategy comes from the offline training stage, where the latent mapping between state space and control policy is built and the time consumption is not critical. Once trained successfully, the DDPG-DRL strategy involves only computationally easy matrix manipulation within the neural networks, which enjoys linear computational complexity. Therefore, in spite of the time-consuming training, the practical implementation of DDPG-DRL strategy is quite tractable.

Alternatively, the absolute CPU time per algorithmic step is a more direct measure of the computational complexity. Tests are hence performed on a laptop with a 2.30 GHz CPU and 16 GB DRAMs. The CPU times for performing the two strategies are shown in TABLE IX. It is shown that the CPU time consumption of the DDPG-DRL strategy is three orders of magnitude lower than that of the MPC controller, suggesting an overriding superiority of the DDPG-DRL strategy in terms of real-time tractability.

It should be noted that a lumped electro-thermal model is used in this paper, and accordingly, only the SOC, polarization voltage and temperature are involved in the state space of MPC controller. Nevertheless, a full consideration of other physical states, like the side reaction potential, solid/liquid phase Li^+ concentration at both electrodes, etc. are demanded for the advanced control of LIB. In this case, a high-order physics model with drastically expanded state space has to be used, and thus the feasibility of MPC controller becomes questionable considering its cubic complexity. In contrast, the DDPG-DRL strategy is expected to still keep an affordable computing cost thanks to its linear complexity. The application of the proposed strategy associated with more complex physics models will be an interesting topic for future investigation.

TABLE IX
CPU TIMES BY USING DIFFERENT STRATEGIES

	DDPG-DRL strategy	MPC-based strategy
CPU time	297.2 μs	195.5 ms

V. CONCLUSION

A DRL-based strategy has been proposed for the thermal- and health- conscious fast charging of LIB. A multi-objective optimization problem has been formulated by penalizing the LIB over-temperature and degradation. Further, an improved environmental perceptive DDPG algorithm with priority experience replay has been exploited to smartly trade-off the charging rapidity and the compliance to physical constraints. The major conclusions are summarized as follows:

(1) The DDPG-DRL strategy validates to fully charge the LIB in 926s without violating the physical constraints.

(2) The CCCV strategy either slows down the charging or recurs the over-heat and quick wear of LIB. Compared to the 6C CCCV strategy, the DDPG-DRL strategy extends the LIB life time by 14.8% with an equivalent charging speed.

(3) The DDPG-DRL strategy performs equivalently with the state-of-the-art MPC controller in the charging rapidity and the compliance to physical constraints. However, the three orders of magnitude lower computational complexity promises a much better potential for real-time utilization.

REFERENCES

[1] Y. Li, B. Xiong, M. Vilathgamuwa, Z. Wei, C. Xie, and C. Zou, "Constrained Ensemble Kalman Filter for Distributed Electrochemical State Estimation of Lithium-Ion Batteries," *IEEE Transactions on Industrial Informatics,* 2020.

[2] Y. Wang, G. Gao, X. Li, and Z. Chen, "A fractional-order model-based state estimation approach for lithium-ion battery and ultra-capacitor hybrid power source system considering load trajectory," *Journal of Power Sources,* vol. 449, p. 227543, 2020.

[3] G. Dong, Z. Chen, J. Wei, and Q. Ling, "Battery Health Prognosis Using Brownian Motion Modeling and Particle Filtering," *IEEE Transactions on Industrial Electronics,* vol. 65, pp. 8646-8655, 2018.

[4] K. Liu, Y. Li, X. Hu, M. Lucu, and D. Widanalage, "Gaussian Process Regression with Automatic Relevance Determination Kernel for Calendar Aging Prediction of Lithium-ion Batteries," *IEEE Transactions on Industrial Informatics,* 2019.

[5] J. Wei, G. Dong, and Z. Chen, "Lyapunov-based thermal fault diagnosis of cylindrical lithium-ion batteries," *IEEE Transactions on Industrial Electronics,* 2019.

[6] X. Hu, K. Zhang, K. Liu, X. Lin, S. Dey, and S. Onori, "Advanced Fault Diagnosis for Lithium-Ion Battery Systems," 2020.

[7] S. S. Zhang, K. Xu, and T. Jow, "Study of the charging process of a LiCoO2-based Li-ion battery," *Journal of power sources,* vol. 160, pp. 1349-1354, 2006.

[8] T. T. Vo, X. Chen, W. Shen, and A. Kapoor, "New charging strategy for lithium-ion batteries based on the integration of Taguchi method and state of charge estimation," *Journal of Power Sources,* vol. 273, pp. 413-422, 2015.

[9] D. Anseán, M. Dubarry, A. Devie, B. Liaw, V. García, J. Viera*, et al.*, "Fast charging technique for high power LiFePO4 batteries: A mechanistic analysis of aging," *Journal of Power Sources,* vol. 321, pp. 201-209, 2016.

[10] P. H. Notten, J. O. het Veld, and J. Van Beek, "Boostcharging Li-ion batteries: A challenging new charging concept," *Journal of Power Sources,* vol. 145, pp. 89-94, 2005.

[11] Y. Li, M. Vilathgamuwa, S. S. Choi, T. W. Farrell, N. T. Tran, and J. Teague, "Development of a degradation-conscious physics-based lithium-ion battery model for use in power system planning studies," *Applied Energy,* vol. 248, pp. 512-525, 2019.

[12] Z. Wei, G. Dong, X. Zhang, J. Pou, Z. Quan, and H. He, "Noise-immune model identification and state of charge estimation for lithium-ion battery using bilinear parameterization," *IEEE Transactions on Industrial Electronics,* 2020.

[13] C. Zhang, J. Jiang, Y. Gao, W. Zhang, Q. Liu, and X. Hu, "Charging optimization in lithium-ion batteries based on temperature rise and charge time," *Applied energy,* vol. 194, pp. 569-577, 2017.

[14] 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.

[15] K. Liu, X. Hu, Z. Yang, Y. Xie, and S. Feng, "Lithium-ion battery charging management considering economic costs of electrical energy loss and battery degradation," *Energy conversion and management,* vol. 195, pp. 167-179, 2019.

[16] S. Pramanik and S. Anwar, "Electrochemical model based charge optimization for lithium-ion batteries," *Journal of Power Sources,* vol. 313, pp. 164-177, 2016.

[17] Z. Chu, X. Feng, L. Lu, J. Li, X. Han, and M. Ouyang, "Non-destructive fast charging algorithm of lithium-ion batteries based on the control-oriented electrochemical model," *Applied energy,* vol. 204, pp. 1240-1250, 2017.

[18] Y. Gao, X. Zhang, B. Guo, C. Zhu, J. Wiedemann, L. Wang*, et al.*, "Health-Aware Multi-objective Optimal Charging Strategy with Coupled Electrochemical-Thermal-Aging Model for Lithium-Ion Battery," *IEEE Transactions on Industrial Informatics,* 2019.

[19] R. Klein, N. A. Chaturvedi, J. Christensen, J. Ahmed, R. Findeisen, and A. Kojic, "Electrochemical Model Based Observer Design for a Lithium-Ion Battery," *IEEE Transactions on Control Systems Technology,* vol. 21, pp. 289-301, 2013.

[20] J. Liu, G. Li, and H. K. Fathy, "An Extended Differential Flatness Approach for the Health-Conscious Nonlinear Model Predictive Control of Lithium-Ion Batteries," *IEEE Transactions on Control Systems Technology,* vol. 25, pp. 1882-1889, 2017.

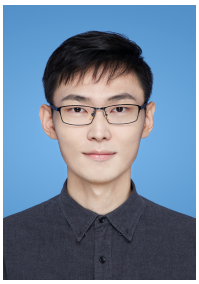
[21] C. Zou, X. Hu, Z. Wei, T. Wik, and B. Egardt, "Electrochemical estimation and control for lithium-ion battery health-aware fast charging," *IEEE Transactions on Industrial Electronics,* vol. 65, pp. 6635-6645, 2017.

[22] L. Zheng, J. Zhu, G. Wang, D. D. Lu, and T. He, "Lithium-ion Battery Instantaneous Available Power Prediction Using Surface Lithium

Concentration of Solid Particles in a Simplified Electrochemical Model," *IEEE Transactions on Power Electronics,* vol. 33, pp. 9551-9560, 2018.
[23] M. A. Xavier and M. S. Trimboli, "Lithium-ion battery cell-level control using constrained model predictive control and equivalent circuit models," *Journal of Power Sources,* vol. 285, pp. 374-384, 2015.
[24] C. Zou, X. Hu, Z. Wei, and X. Tang, "Electrothermal dynamics-conscious lithium-ion battery cell-level charging management via state-monitored predictive control," *Energy,* vol. 141, pp. 250-259, 2017.
[25] H. Fang, Y. Wang, and J. Chen, "Health-aware and user-involved battery charging management for electric vehicles: Linear quadratic strategies," *IEEE Transactions on Control Systems Technology,* vol. 25, pp. 911-923, 2016.
[26] Q. Ouyang, Z. Wang, K. Liu, G. Xu, and Y. Li, "Optimal Charging Control for Lithium-Ion Battery Packs: A Distributed Average Tracking Approach," *IEEE Transactions on Industrial Informatics,* 2019.
[27] Q. Ouyang, J. Chen, J. Zheng, and H. Fang, "Optimal multiobjective charging for lithium-ion battery packs: A hierarchical control approach," *IEEE Transactions on Industrial Informatics,* vol. 14, pp. 4243-4253, 2018.
[28] N. Tian, H. Fang, and Y. Wang, "Real-Time Optimal Lithium-Ion Battery Charging Based on Explicit Model Predictive Control," *IEEE Transactions on Industrial Informatics,* pp. 1-1, 2020.
[29] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*: MIT press, 2018.
[30] A. Chiş, J. Lundén, and V. Koivunen, "Reinforcement learning-based plug-in electric vehicle charging with forecasted price," *IEEE Transactions on Vehicular Technology,* vol. 66, pp. 3674-3684, 2016.
[31] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using q-learningtechnique," *IEEE Transactions on Industrial Informatics,* 2020.
[32] 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, pp. 4513-4521, 2020.
[33] X. Lin, H. E. Perez, S. Mohan, J. B. Siegel, A. G. Stefanopoulou, Y. Ding, *et al.*, "A lumped-parameter electro-thermal model for cylindrical batteries," *Journal of Power Sources,* vol. 257, pp. 1-11, 2014.
[34] S. Ebbesen, P. Elbert, and L. Guzzella, "Battery state-of-health perceptive energy management for hybrid electric vehicles," *IEEE Transactions on Vehicular technology,* vol. 61, pp. 2893-2900, 2012.
[35] Z. Wei, J. Zhao, D. Ji, and K. J. Tseng, "A multi-timescale estimator for battery state of charge and capacity dual estimation based on an online identified model," *Applied energy,* vol. 204, pp. 1264-1274, 2017.

Zhongbao Wei (M'19–SM'21) received the B.Eng. and the M.Sc. degrees in instrumental science and technology from Beihang University, China, in 2010 and 2013, and the Ph.D. degree in power engineering from Nanyang Technological University, Singapore, in 2017.
He has been a research fellow with Energy Research Institute @ NTU, Nanyang Technological University from 2016 to 2018. He is currently a Professor in vehicle engineering with the National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, China. He has authored more than 60 peer-reviewed articles. His research interests include battery management and energy management for hybrid energy systems. He serves as an Associate Editor for IET renewable power generation, IET intelligent transportation, and a Guest Editor for IEEE Journal of Emerging and Selected Topics in Power Electronics.

Zhongyi Quan (S'12-M'19) received the B.Eng. degree in instrumental science from Tianjin University, Tianjin, China in 2010, M.Sc. degree in instrumental science from Beihang University, Beijing, China, in 2013, and Ph.D. degree in electrical engineering in 2019 from University of Alberta, Canada, where he is currently a Postdoctoral Fellow. He is the recipient of IEEE PELS ECCE Best Student Project Demonstration on Emerging Technology 3rd Prize Award in 2018 and the Alberta GreenSTEM Fellowship. His current research interests include design of high density and high efficiency power converters for applications such as renewable energy, electric vehicles, and other energy efficient equipment.



Jingda Wu received his B.S. (2016) in vehicle engineering and M.Sc. (2019) in mechanical engineering from Beijing Institute of Technology, China.
He is currently working on his Ph.D. in mechanical engineering at Nanyang Technological University, Singapore. His research mainly focuses on optimization and control of human-machine collaborated driving, design of autonomous driving strategy with machine learning methods, energy management of electric vehicle and Li-ion battery.

Yang Li (S'11–M'16) received the B.E. degree in electrical engineering from Wuhan University, Wuhan, China, in 2007, and the M.Sc. and Ph.D. degrees in power engineering from the Nanyang Technological University (NTU), Singapore, in 2008 and 2015, respectively. From 2016 to 2018, he was a Research Fellow with the School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia. Since 2019, he has been an Associate Professor with the School of Automation, Wuhan University of Technology, Wuhan, China. He is currently a Researcher with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. His research interests include modeling, control, and application of renewable and energy storage systems in power grid and transport sectors. Dr. Li was a recipient of the EU Marie Skłodowska-Curie Individual Fellowship in 2020.

Josep Pou (S'97–M'03–SM'13–F'17) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Technical University of Catalonia (UPC)-Barcelona Tech, in 1989, 1996, and 2002, respectively.
In 1990, he joined the faculty of UPC as an Assistant Professor, where he became an Associate Professor in 1993. From February 2013 to August 2016, he was a Professor with the University of New South Wales (UNSW), Sydney, Australia. He is currently a Professor with the Nanyang Technological University (NTU), Singapore, where he is Program Director of Power Electronics at the Energy Research Institute at NTU (ERI@N) and co-Director of the Rolls-Royce at NTU Corporate Lab. He has authored more than 350 published technical papers and has been involved in several industrial projects and educational programs in the fields of power electronics and systems. His research interests include modulation and control of power converters, multilevel converters, renewable energy, energy storage, power quality, HVdc transmission systems, and more-electrical aircraft and vessels.
He is currently Associate Editor of the IEEE Journal of Emerging and Selected Topics in Power Electronics. He was co-Editor-in-Chief and Associate Editor of the IEEE Transactions on Industrial Electronics. He received the 2018 IEEE Bimal Bose Award for Industrial Electronics Applications in Energy Systems.

Hao Zhong received a bachelor's degree in vehicle engineering from Beijing Institute of technology in 2020. He is currently working toward the Ph.D. degree at school of mechanical engineering, Beijing Institute of technology. His current research interests include modeling and control of energy storage system.