Physics-Informed Nonlinear Extension Techniques for Robust Joint State Estimation of Li-Ion Batteries

Peng Guo, Xinghua Liu, *Senior Member, IEEE*, Wentao Ma, *Member, IEEE*, Yang Li, *Senior Member, IEEE*, Gaoxi Xiao, *Senior Member, IEEE*, Zhongbao Wei, *Senior Member, IEEE*, and Badong Chen, *Senior Member, IEEE*

***Abstract*—To address the challenges of poor noise immunity and limited generalization performance in Li-ion battery modeling and state estimation (SE), a novel robust framework for parameter identification (PI) and joint estimation of state of charge (SOC) and surface temperature is proposed in this study by leveraging physical information and nonlinear extension techniques. Initially, a robust forgetting factor recursive maximum total correntropy algorithm is developed for PI, providing a solid foundation for SE under noisy conditions. Subsequently, a robust SOC estimation method is formulated by embedding the maximum correntropy criterion (MCC) with an adaptive kernel width into the square-root cubature Kalman filter, effectively replacing the conventional mean square error with MCC to enhance noise resilience. Next, a multidimensional feature input set is constructed using the PI results, including total internal resistance as auxiliary physical information, along with SOC estimates and raw measurements. A subinput structure is further designed using partial correlation analysis, and then the extreme learning machines are utilized to project the subinputs into a high-dimensional (HD) feature space to extract latent correlation features. Finally, by integrating nonlinear extended features with raw data in parallel, the input to the bidirectional gated recurrent unit model is generated, enabling simultaneous extraction of global representations from both HD and low-dimensional spaces. Experimental results demonstrate that the proposed method outperforms existing advanced approaches in SE under strong noise interference and complex operational conditions.**

***Index Terms*—Bidirectional gated recurrent unit (GRU), joint state estimation, Li-ion battery (LIB), nonlinear extension, parameter identification, physical information.**

I. Introduction

LI-ION batteries (LIBs) have emerged as the preferred power source for energy storage systems (ESSs) and electric vehicles (EVs) due to their high power density, high energy density, and cost-effectiveness [1], [2]. The safe and stable operation of LIBs relies on the performance and reliability of the battery management system. Among the key parameters it monitors, state of charge (SOC) and state of temperature are critical for maintaining operational stability and preventing hazardous conditions. Accurate estimation of these parameters ensures reliable driving range predictions, slows LIB degradation, and reduces the risk of thermal runaway [3], [4]. Theoretically, temperature sensors can directly measure the surface temperature (ST) and core temperature of LIBs. In practice, ESSs and EVs contain hundreds or thousands of cells, making direct measurement impractical [5], and temperature is often correlated with the SOC. Additionally, LIBs may encounter electromagnetic interference, sensor failures, and other disturbances during operation, leading to data contamination by Gaussian (GN) or non-Gaussian noise (NGN) [6], [7]. Thus, developing an adaptive joint estimation framework for SOC and ST is crucial considering noisy measurement.

Contemporary ST estimation methodologies for LIB ST are systematically categorized into three types: 1) electrochemical impedance spectroscopy (EIS) [8]; 2) model-based [9]; and 3) data-driven [10] approaches. The EIS method quantifies temperature-impedance correlations by analyzing voltage responses to controlled alternating current excitation at defined frequencies. However, this process

Received 15 August 2025; revised 2 December 2025; accepted 31 December 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62473308 and Grant 62473309, in part by the Shaanxi Outstanding Youth Science Fund Project under Grant 2024JC-JCQN-68, in part by the National Key R&D Program of China under Grant 2021YFB2401900, in part by the Joint Fund Project of the National Natural Science Foundation of China under Grant U21A20485, in part by the Xi'an Science and Technology Plan Project under Grant 23GXFW0071 and Grant 24GXFW0050, and in part by the Xi'an Key Laboratory under Grant 24ZDSY0015. *(Corresponding authors: Xinghua Liu; Wentao Ma.)*

Peng Guo, Xinghua Liu, and Wentao Ma are with the School of Electrical Engineering, Xi'an University of Technology, Xi'an 710048, China (e-mail: 2211921110@stu.xaut.edu.cn; liuxh@xaut.edu.cn; mawt@xaut.edu.cn).

Yang Li is with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden (e-mail: yangli@ieee.org).

Gaoxi Xiao is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: egxxiao@ntu.edu.sg).

Zhongbao Wei is with the National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China (e-mail: weizb@bit.edu.cn).

Badong Chen is with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710048, China (e-mail: chenbd@mail.xjtu.edu.cn).

Digital Object Identifier 10.1109/TIE.2026.3651401

needs specialized instrumentation and strictly controlled steady-state operating conditions to mitigate polarization interference [9], [10], substantially limiting its applicability. The model-based method employs an electrothermal coupling model that integrates electrical and thermal models. When combined with Kalman filter (KF) algorithms [11], these methods exhibit enhanced capability in capturing the intrinsic heat generation mechanisms of LIBs while maintaining strong interpretability. However, their estimation accuracy is fundamentally constrained by the battery's heat generation rate, thermal properties, and boundary conditions.

Unlike EIS and model-based approaches, the neural network (NN) is an outstanding data-driven approach, which can establish a canonical state estimation (SE) model to effectively capture the complex nonlinear dynamics of LIBs. NNs primarily encompass two categories: 1) feedforward NNs (FNNs); and 2) recurrent NNs (RNNs) [12], [13], [14]. Among these, the extreme learning machine (ELM), as a type of FNN, has been applied for the SE of LIBs [13]. It is renowned for its rapid training capability and the nonlinear mapping characteristics of its hidden layer. Specifically, it can directly map input data into a high-dimensional (HD) feature space through random initialization of hidden layer parameters, thereby effectively addressing complex nonlinear problems [14]. However, existing research has overlooked this characteristic of ELM, and FNNs perform poorly at processing temporal information and dynamic data, limiting their application in SE of LIB. In contrast, RNNs are capable of effectively handling sequential data and demonstrate significant advantages in dynamic data modeling, particularly with their improved models such as long short-term memory (LSTM) and gated recurrent units (GRU). These models successfully resolve the issues of vanishing and exploding gradients in traditional RNNs during long sequence data processing, significantly enhancing the model's ability to model long-term dependencies [14], [15]. Yao et al. [10] developed a high-precision ST estimation method based on a GRU-RNN framework, incorporating distribution analysis and adaptive data normalization across varying temperatures. Although it demonstrated strong intradataset consistency, validation was limited to identical dynamic conditions, leaving its dynamic generalization, particularly under significant temperature fluctuations, unverified. Zheng et al. [5] proposed an improved convolutional NN (CNN) LSTM (ICNN-LSTM) framework that integrates convolutional feature extraction with electrothermal model inputs, resulting in enhanced estimation accuracy. However, this method imposes prohibitive computational costs for parameter identification (PI) in centralized electrothermal models, while fixed-parameter assumptions fail to capture the transient behaviors of LIBs under dynamic conditions. Moreover, the approaches discussed above typically depend on high-precision battery testers to obtain SOC via ampere-hour integration, which is then used as input for neural networks. In practice, however, SOC cannot be directly measured, necessitating further verification of estimation reliability [4], [9]. Additionally, LSTM and GRU architectures process information only in the forward direction, limiting their ability to fully capture contextual dependencies in battery thermal sequences.

To overcome this critical limitation, bidirectional GRU (Bi-GRU) introduces parallel backward temporal analysis, enabling systematic capture of latent electrochemical-thermal interactions through simultaneous forward-backward sequence processing, particularly improving SE accuracy under rapid thermal transients [16], [17], [18]. Although the Bi-GRU network demonstrates enhanced temporal modeling through dual-sequence analysis, its feature information extraction capability is restrained under data scarcity. In practice, most data-driven methods must rely on very limited measurement data for model training, often overlooking the physical characteristics of LIBs and the impact of HD data features on SE [5], [14]. However, physical characteristics are essential for capturing the thermal and degradation behaviors of LIBs, while HD features are critical for modeling their multidimensional dynamics. Neglecting these factors compromises the model's generalization and robustness. Thus, effectively integrating physical characteristics with HD nonlinear feature extraction to enhance ST estimation accuracy remains a critical challenge.

An effective solution to the above problem is to capture the physical characteristics of LIBs through electrical models (EMs), which form the foundation for model-based SOC estimation methods. This physical information is integrated with SOC estimates, serving as supplementary input for ST estimation. Recent studies on SOC estimation often use voltage, current, and ST as primary inputs. However, directly obtaining ST in practical applications is challenging, which significantly complicates data-driven SOC estimation methods [6], [19]. In contrast, model-based SOC estimation methods are more mature, enabling high-precision battery modeling through the construction of concise EMs. Moreover, PI algorithms are straightforward to implement, and incorporating KF algorithms or observer techniques can significantly enhance the accuracy of SOC estimation [6], [19], [20]. Nonetheless, existing studies often overlook the impacts of electromagnetic interference and data transmission errors, which can introduce GN or NGN into measurement data [6], [7], [21]. To enhance the robustness of SOC estimation under noisy conditions, recent studies have incorporated the maximum correntropy criterion (MCC) [22], [23] to improve the performance of KF methods in complex noise environments. However, research on the selection and optimization of the kernel width (Kw) parameter remains limited, constraining the performance of enhanced KF methods in practical applications [21], [24]. In addition, robust PI design under noise interference has yet to be effectively addressed, and the influence of SOC estimation on ST accuracy under dynamic conditions remains insufficiently explored. Further theoretical refinement and experimental validation are needed in these areas. To address the aforementioned challenges, a robust framework for PI and joint SE is proposed. The primary contributions of this work are summarized as follows.

1) A robust method for PI and SOC estimation, integrating correntropy theory with a Kw adaptation (KwA) mechanism, is proposed, markedly improving estimation accuracy under complex noise conditions.
2) Partial correlation analysis (PCA) is employed to uncover nonlinear couplings among multidimensional inputs and

the ST, thereby informing the design of subinput structures that retain intervariable dependencies. These subinputs are then projected into a HD feature space via the ELM hidden layer, whereby key nonlinear features are efficiently extracted.

3) An enhanced Bi-GRU model, termed PINE-Bi-GRU, is proposed, integrating physical information and nonlinear expansion. By concurrently fusing nonlinear features with raw data as inputs, the model enables effective capture of global features across both HD and low-dimensional (LD) spaces.

The remainder of this article is organized as follows. Section II reviews the fundamentals of battery modeling, PI, and SOC estimation. Section III details the design and implementation of the PINE-Bi-GRU framework. Section IV discusses the experimental results and comprehensive evaluations. Finally, Section V concludes the study and outlines future research directions.

II. Methods for PI and SOC Estimation

A. Electrical Model of the LIB

In LIB systems, the SOC and ST are generally challenging to measure directly. Consequently, indirect estimation is pursued through in-depth analysis and feature extraction, utilizing limited data from voltage, current, and ambient temperature (AT). This study focuses on the extraction of additional feature signals from constrained datasets and the achievement of accurate SE, which are identified as the central research challenges. To address this challenge, a first-order RC equivalent circuit model (RC-ECM) is employed as the foundational framework. This model is characterized by its structural simplicity and ability to effectively capture the essential dynamic characteristics of battery systems. Precise estimation of ECM parameters and SOC is achieved by integrating robust PI methods with the KF. As illustrated in Fig. 1, the first-order RC-ECM consists of the open-circuit voltage (OCV) V_{oc}, which is correlated with the SOC, the ohmic resistance R_o, polarization resistance R_p, and polarization capacitance C_p. Based on Kirchhoff's laws, the electrical behavior of the first-order RC-ECM is described as

$$\begin{cases} V_p/R_p + C_p dV_p/dt = I \\ V_t = V_{oc} - V_p - R_o I \end{cases} \tag{1}$$

where I represents the load current, with discharging defined as positive and charging as negative. The polarization voltage and terminal voltage of the battery are denoted as V_p and V_t, respectively. Based on (1), a transfer function can be derived

$$d(s)/I(s) = -[R_o + R_p/(1 + R_p C_p s)] \tag{2}$$

where $d = V_t - V_{oc}$. Next, applying the bilinear transformation $s = 2(q-1)/\Delta t/(q+1)$ to (2) yields

$$d(q^{-1})/I(q^{-1}) = (a_0 + a_1 q^{-1})/(1 + a_2 q^{-1}) \tag{3}$$

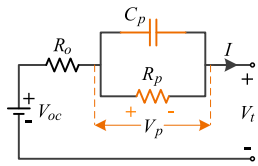

Fig. 1. Structure of the first-order RC-ECM.

where Δt is the sampling time and q represents the discrete operator. Therefore, the coefficients in (3) can be expressed as

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -(\Delta t R_o + \Delta t R_p + 2R_p C_p R_o)/(\Delta t + 2R_p C_p) \\ -(\Delta t R_o + \Delta t R_p - 2R_p C_p R_o)/(\Delta t + 2R_p C_p) \\ (\Delta t - 2R_p C_p)/(\Delta t + 2R_p C_p) \end{bmatrix}. \tag{4}$$

Based on the derived discrete-time system (3), a regression model can be established to facilitate numerical analysis

$$d_t = \mathbf{w}_t^T \mathbf{x}_t \tag{5}$$

where $\mathbf{x}_t = \begin{bmatrix} -d_{t-1} & I_t & I_{t-1} \end{bmatrix}^T$, $\mathbf{w}_t = \begin{bmatrix} a_{2,t} & a_{0,t} & a_{1,t} \end{bmatrix}^T$, and T denotes the transpose operation. The model parameters are derived through backward inference by using (4) once the regression model (5) is solved

$$\begin{bmatrix} C_p & R_p & R_o \end{bmatrix} = \begin{bmatrix} \frac{(1-a_2)^2}{4(a_0 a_2 - a_1)}\Delta t & 2\frac{a_0 a_2 - a_1}{1-a_2^2} & \frac{a_1 - a_0}{1-a_2} \end{bmatrix}. \tag{6}$$

The total internal resistance of LIB serves as a critical physical indicator for characterizing dynamic battery behavior and enabling temperature estimation, which is defined as

$$R_T = R_p + R_o. \tag{7}$$

The R_T is inherently sensitive to electrochemical-thermal coupling mechanisms. First, R_T demonstrates a well-established monotonic relationship with temperature, governed by Arrhenius-type kinetics that influence ionic conductivity in the electrolyte and charge transfer reactions at the electrodes. Specifically, R_o decreases exponentially with increasing temperature due to enhanced ion mobility, while R_p reflects kinetic limitations that scale inversely with thermal activation energy. The dual dependence of R_T allows it to capture real-time thermal state information. Second, as a lumped parameter in ECMs, R_T captures transient voltage-current responses during operational load changes, such as charge and discharge pulses. These dynamics are influenced by temperature-dependent diffusion coefficients and interfacial reaction rates, positioning R_T as a comprehensive descriptor of electrochemical-thermal interactions.

Remark 1: Aging effects gradually alter internal resistance through reduced electrode activity, solid electrolyte interphase growth, and electrolyte side reactions, which introduce bias into PI and degrade SE accuracy. However, such resistance drift evolves slowly over many cycles compared with the rapid perturbations induced by temperature and load [5].

B. SOC Estimation Based on Improved MCC Square-Root Cubature KF

Given the harsh operating conditions of batteries, which are susceptible to external factors such as electromagnetic interference and communication failures, two primary design challenges are encountered by existing KF methods [21], [24]. First, conventional KF methods based on the mean squared error criterion are effective under GN conditions but inadequate for accurate estimation under complex real-world noise disturbances [21]. Second, robust KF methods utilizing the MCC or mixture MCC provide precise SOC estimation under complex noise scenarios. However, they introduce high structural complexity, leading to an increased number of free parameters and heightened challenges in parameter optimiz-ation [24]. Therefore, the development of a robust filtering algorithm, characterized by structural simplicity and strong noise immunity, is essential. In this study, the square-root cubature KF (SRCKF) with MCC (called MCC-SRCKF) is introduced to perform noise-immune SOC estimation. Furthermore, an adjustment mechanism is incorporated to dynamically optimize the key parameter Kw in MCC, improving the convergence speed and robustness of MCC. For brevity, only the derivation of the MCC framework with the KwA mechanism is presented in this section.

Correntropy is defined as the expected value of the kernel function between two random variables, B_1 and B_2 [23], [25]

$$\mathrm{V}(B_1, B_2) = \mathrm{E}[\kappa_\sigma(B_1 - B_2)] \tag{8}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator, and $\kappa_\sigma(\cdot)$ refers to the Gaussian kernel function with a Kw $\sigma > 0$, which are specifically defined as

$$\kappa_\sigma(B_1 - B_2) = G_\sigma(B_1 - B_2) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-|B_1 - B_2|^2/2\sigma^2) \tag{9}$$

where σ is the Kw. In practical applications, the applicability of (8) is limited by uncertainties in error distribution and the limited availability of sample data. To address these limitations, we redefine (8) as

$$\hat{\mathrm{V}}(B_1, B_2) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} \exp(-|b_{1,t} - b_{2,t}|^2/2\sigma^2) \right) \tag{10}$$

where $b_{1,t}$ and $b_{2,t}$ are the tth components of B_1 and B_2, respectively, and n denotes the sample size. In practical applications, B_1 and B_2 are associated with the true and the estimated values of the measured information, respectively, with discrepancies between them typically attributed to interference from unknown noise [25]. In the domain of information-theoretic learning, the MCC is widely recognized as a standard optimization metric and is formally defined as

$$J_{\mathrm{MCC}} = \max \hat{\mathrm{V}}(B_1, B_2). \tag{11}$$

Evidently, only a single free parameter σ is used in the MCC model. Building on this simplicity, an efficient KwA mechanism (12) is developed to track the dynamic variation of error information, enabling real-time optimization of Kw during the algorithm's iterative process

$$\sigma = \max\left\{2\sigma_o - (2\sigma_o - 1)^{\exp(-|e_t|^2/2)}, \sigma_o\right\} \tag{12}$$

where $e_t = b_{1,t} - b_{2,t}$ and $\sigma_o(\sigma_o > 1)$ is defined as the initial value of Kw. According to (12), the exponential term is constrained within the interval [0, 1], and the range of σ can be derived as

$$\sigma_o \le \sigma < 2\sigma_o - 1. \tag{13}$$

During the SOC estimation process, significant errors are often introduced due to challenges in accurately determining the initial SOC. Moreover, noise in battery data inevitably degrades the convergence rate of the MCC-SRCKF algorithm. These problems are addressed by the proposed KwA mechanism, which dynamically adjusts the parameter Kw to enhance algorithmic performance. In the initial iterations, the KwA mechanism typically selects a larger parameter value (e.g., $2\sigma_o - 1$) to expedite convergence. As iterations proceed and estimation errors decrease, Kw is adaptively tuned based on the error's dynamic behavior, with smaller σ gradually adopted, thereby significantly improving the algorithm's robustness. Ultimately, the KwA mechanism effectively captures variations in estimation error and optimizes parameter settings in real-time, leading to enhanced accuracy and robustness of SOC estimation when the improved MCC-SRCKF (IMCC-SRCKF) algorithm is employed.

C. Parameter Identification Based on Improved Forgetting Factor Recursive Maximum Total Correntropy (IFRMTC)

In dynamic environments with varying temperatures, long-term battery operation is susceptible to electromagnetic interference and communication errors, which introduce both GN and NGN into the collected data [6], [7], [21]. To address these challenges, we propose a robust recursive maximum total correntropy (RMTC) algorithm for PI. The algorithm incorporates both input and output noise-induced errors into the cost function, capturing higher-order error statistics to enhance PI accuracy. A forgetting factor is introduced to develop the FRMTC algorithm, which prioritizes recent data to mitigate estimation lag and more accurately reflect the battery's dynamic behavior. Additionally, integrating a KwA mechanism results in the IFRMTC, which updates Kw dynamically by synthesizing total error information with the weight vector from previous iterations. This allows for real-time tracking of noise characteristics, accelerating convergence and improving robustness in complex noise environments. The detailed methodology is outlined as follows.

To mitigate the interference induced by input and output noises, the classical errors-in-variables model is employed in this study to characterize this issue

$$\begin{cases} \hat{\mathbf{x}}_{1,t} = \mathbf{x}_{1,t} + \mathbf{u}_t \\ \hat{d}_t = d_t + v_t. \end{cases} \tag{14}$$

In this context, $\mathbf{x}_{1,t}$ and d_t are defined as the input and output vectors of the system, respectively, while $\mathbf{u}_t$ represents the input noise with a covariance of $\varsigma_i^2\mathbf{I}_k$, and v_t denotes the output noise with a covariance of $\varsigma_o^2\mathbf{I}_k$. The identity matrix is indicated by $\mathbf{I}_k$, and k represents the filter order. In addition, $\hat{\mathbf{x}}_{1,t}$ and $\hat{d}_t$ are the input and output vectors of the system that are influenced by noise. Furthermore, it is assumed that the input noise and output noise are statistically independent and follow the same type of distribution.

Inspired by the MCC [22], [23], the optimal weight $\mathbf{w}$ can be determined by maximizing the cost function J_{MTC}

$$\begin{cases} J_{\text{MTC}} = \frac{1}{n}\sum_{t=1}^{n}\lambda^{n-t}(\ell(e_t)/(\sqrt{2\pi}\sigma_{1,t})) \\ \ell(e_t) = \exp(-|e_t|^2/(2\sigma_{1,t}^2\|\hat{\mathbf{w}}_t\|_2^2)). \end{cases} \tag{15}$$

The parameter λ $(0.95 < \lambda < 1)$ represents the forgetting factor, which is set to 0.997 in this study. The error is denoted as $e_t = \hat{d}_t - \mathbf{w}^T\hat{\mathbf{x}}_{1,t}$, where $\mathbf{w}$ represents the filter weight. The total error $e_t/\sqrt{\hat{\mathbf{w}}^T\hat{\mathbf{w}}}$ is caused by the combined effects of input noise and output noise, with $\hat{\mathbf{w}} = \begin{bmatrix}\sqrt{\beta} & -\mathbf{w}^T\end{bmatrix}^T$, $\beta = \varsigma_o^2/\varsigma_i^2$. Furthermore, σ_1 represents the Kw associated with the J_{MTC}. By combining (5) and (14), it can be concluded that the total error is primarily influenced by input and output noises [25]. The fixed-point solution of the filtering weights is obtained by solving the gradient of the objective function $\partial J_{\text{MTC}}/\partial\mathbf{w}_t = 0$, where

$$\mathbf{w}_t = \left\{\sum_{t=1}^{n}\lambda^{n-t}\left[\ell(e_t)(\hat{\mathbf{x}}_{1,t}\hat{\mathbf{x}}_{1,t}^T\|\hat{\mathbf{w}}_t\|_2^2 - |e_t|^2\mathbf{I}_k)\right]\right\}^{-1}\sum_{t=1}^{n}\lambda^{n-t}[\ell(e_t)\hat{d}_t\hat{\mathbf{x}}_{1,t}]\|\hat{\mathbf{w}}_t\|_2^2. \tag{16}$$

Here, $\mathbf{w}$ is dynamically adjusted in real time based on a limited sample set, and it follows:

$$\mathbf{w}_t = (\mathbf{R}_t^*)^{-1}\hat{\mathbf{p}}_t\|\hat{\mathbf{w}}_{t-1}\|_2^2 \tag{17}$$

where $\mathbf{R}_t^* = \hat{\mathbf{R}}_{1,t}\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \hat{\mathbf{R}}_{2,t}$

$$\hat{\mathbf{R}}_{1,t} = \sum_{l=1}^{t}\lambda^{t-l}[\ell_1(e_l)\hat{\mathbf{x}}_{1,l}\hat{\mathbf{x}}_{1,l}^T] \tag{18}$$

$$\hat{\mathbf{R}}_{2,t} = -\sum_{l=1}^{t}\lambda^{t-l}[\ell_1(e_l)|e_l|^2]\mathbf{I}_k \triangleq \alpha_t\mathbf{I}_k \tag{19}$$

$$\hat{\mathbf{p}}_t = \sum_{l=1}^{t}\lambda^{t-l}[\ell_1(e_l)\hat{d}_l\hat{\mathbf{x}}_{1,l}] \tag{20}$$

where $\ell_1(e_l) = \exp(-|e_l|^2/(2\sigma_{1,t}^2\|\hat{\mathbf{w}}_{l-1}\|_2^2))$.Then, the update equations of the $\hat{\mathbf{R}}_{1,t}$,$\hat{\mathbf{R}}_{2,t}$ and $\hat{\mathbf{p}}_t$ are obtained as

$$\hat{\mathbf{R}}_{1,t} = \lambda\hat{\mathbf{R}}_{1,t-1} + \ell_1(e_t)\hat{\mathbf{x}}_{1,t}\hat{\mathbf{x}}_{1,t}^T \tag{21}$$

$$\hat{\mathbf{R}}_{2,t} = \lambda\hat{\mathbf{R}}_{2,t-1} - \ell_1(e_t)|e_t|^2\mathbf{I}_k \triangleq \alpha_{t-1}\mathbf{I}_k - \gamma_t\mathbf{I}_k \tag{22}$$

$$\hat{\mathbf{p}}_t = \lambda\hat{\mathbf{p}}_{t-1} + \ell_1(e_t)\hat{d}_t\hat{\mathbf{x}}_{1,t} \tag{23}$$

where $\ell_1(e_t) = \exp(-|e_t|^2/(2\sigma_{1,t}^2\|\hat{\mathbf{w}}_{t-1}\|_2^2))$.

Remark 2: Since J_{MTC} is not a strictly convex function, the weight $\mathbf{w}$ is updated using (24) during iterations before $t = t_o$ to ensure the positive definiteness of $\mathbf{R}_t^*$. The value of t_o is determined through iterative experimental optimization

$$\mathbf{w}_t = (\hat{\mathbf{R}}_{1,t}\|\hat{\mathbf{w}}_{t-1}\|_2^2)^{-1}\hat{\mathbf{p}}_t\|\hat{\mathbf{w}}_{t-1}\|_2^2. \tag{24}$$

Each iteration of (17) requires matrix inversion, which substantially increases computational complexity and time overhead. To mitigate this, the matrix inversion lemma (MIL) and associated computational techniques are employed [25]. As a result, matrix $(\mathbf{R}_t^*)^{-1}$ is reformulated as

$$(\mathbf{R}_t^*)^{-1} = \hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 - \left[(\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)(\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \hat{\mathbf{R}}_{2,t}^{-1})^{-1}\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2\right]. \tag{25}$$

Combining (17) and (25) produces

$$\mathbf{w}_t = \hat{\mathbf{R}}_{1,t}^{-1}\hat{\mathbf{p}}_t - [(\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)(\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \hat{\mathbf{R}}_{2,t}^{-1})^{-1}\hat{\mathbf{R}}_{1,t}^{-1}\hat{\mathbf{p}}_t] = \mathbf{w}_t^* - \mathbf{P}_t\mathbf{w}_t^* \tag{26}$$

where $\mathbf{P}_t = (\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)(\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \hat{\mathbf{R}}_{2,t}^{-1})^{-1}$ and $\mathbf{w}_t^* = \hat{\mathbf{R}}_{1,t}^{-1}\hat{\mathbf{p}}_t$. Equation (25) can be transformed into the following form using the MIL method:

$$\hat{\mathbf{R}}_{1,t}^{-1} = \lambda^{-1}\hat{\mathbf{R}}_{1,t-1}^{-1} - \mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1} \tag{27}$$

$$\mathbf{b}_t = \lambda^{-1}\hat{\mathbf{R}}_{1,t-1}^{-1}\hat{\mathbf{x}}_{1,t}(\lambda(\ell_1(e_t))^{-1} + \hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1}\hat{\mathbf{x}}_{1,t})^{-1}. \tag{28}$$

Equation (28) can be simplified by multiplying both sides by $\lambda + \ell_1(e_t)\hat{\mathbf{x}}_t^T\hat{\mathbf{R}}_{1,t-1}^{-1}\hat{\mathbf{x}}_t$, i.e.,

$$\mathbf{b}_t = \lambda^{-1}\left[\ell_1(e_t)(\lambda^{-1}\hat{\mathbf{R}}_{1,t-1}^{-1} - \mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1})\hat{\mathbf{x}}_{1,t}\right] = \lambda^{-1}\ell_1(e_t)\hat{\mathbf{R}}_{1,t}^{-1}\hat{\mathbf{x}}_{1,t}. \tag{29}$$

Based on (23), (27), and (29), a recursive expression for updating $\mathbf{w}_t^*$ is derived as

$$\mathbf{w}_t^* = \hat{\mathbf{R}}_{1,t}^{-1}\hat{\mathbf{p}}_t \approx \hat{d}_t\mathbf{b}_t + (\lambda^{-1}\hat{\mathbf{R}}_{1,t-1}^{-1} - \mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1})\lambda\hat{\mathbf{p}}_{t-1} = \hat{d}_t\mathbf{b}_t + \hat{\mathbf{R}}_{1,t-1}^{-1}\hat{\mathbf{p}}_{t-1} - \lambda\mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1}\hat{\mathbf{p}}_{t-1} = \mathbf{w}_{t-1}^* + \mathbf{b}_t e_t^* \tag{30}$$

$$e_t^* = \hat{d}_t - \lambda\hat{\mathbf{x}}_{1,t}^T\mathbf{w}_{t-1}^*. \tag{31}$$

It should be noted that matrix inversion is required to calculate $\mathbf{P}_t$. To reduce computational load, a transformation based on the MIL is applied to approximate $\mathbf{P}_t$. Here, an auxiliary variable $\mathbf{T}_t$ is introduced to facilitate this process

$$\mathbf{T}_t = \hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \hat{\mathbf{R}}_{2,t}^{-1}. \tag{32}$$

Using (22) and (27), (32) can be reformulated as

$$\begin{aligned}\mathbf{T}_t &= \lambda^{-1}\hat{\mathbf{R}}_{1,t-1}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 - \mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + 1/\alpha_{t-1}\mathbf{I}_k - 1/\alpha_{t-1}\mathbf{I}_k + 1/\alpha_t\mathbf{I}_k \\ &\approx \lambda^{-1}\mathbf{T}_{t-1} - \mathbf{b}_t\hat{\mathbf{x}}_{1,t}^T\hat{\mathbf{R}}_{1,t-1}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2 + \gamma_t/[(\alpha_{t-1}-\gamma_t)\alpha_{t-1}]\mathbf{I}_k.\end{aligned} \tag{33}$$

Then, let $\mathbf{Z}_t = \{\lambda^{-1}\mathbf{T}_{t-1} + \gamma_t/[(\alpha_{t-1}-\gamma_t)\alpha_{t-1}]\mathbf{I}_k\}^{-1}$. Applying the MIL theory yields an alternative form of $\mathbf{Z}_t$ and $\mathbf{T}_t^{-1}$ as

$$\mathbf{Z}_t = \lambda\mathbf{T}_{t-1}^{-1} - \lambda\mathbf{T}_{t-1}^{-1}[(((\alpha_{t-1}-\gamma_t)\alpha_{t-1})/\gamma_t)\mathbf{I}_k$$

Algorithm 1: The PI Algorithm Utilizing the IFRMTC Framework.

Input: $\hat{d}_t, \hat{\mathbf{x}}_{1,t}$.
Output: $\mathbf{w}_t$.
Parameter settings: $\sigma_{1,t}, \alpha, t_o$.
1. Initializations: set $\mathbf{w}_o = \mathbf{w}_o^* = \alpha_o = \varepsilon, \hat{\mathbf{R}}_{1,o} = \varepsilon \mathbf{I}_m$. (where ε denotes the regularization parameter.)
2. For $t = 1$ to M
3. Compute: $e_t = \hat{d}_t - \mathbf{w}_{t-1}^T \hat{\mathbf{x}}_t$. $o(k)$
4. Compute: $e_t^* = \hat{d}_t - (\mathbf{w}_{t-1}^*)^T \hat{\mathbf{x}}_t$. $o(k)$
5. Compute: $\|\hat{\mathbf{w}}_{t-1}\|_2^2 = \mathbf{w}_{t-1}^T \mathbf{w}_{t-1} + \beta$. $o(k)$
6. Compute: σ_t using Eq. (12). $o(k)$
7. Compute: $\mathbf{b}_t$ using Eq. (28). $o(k^2)$
8. Compute: $\mathbf{w}_t^* = \mathbf{w}_{t-1}^* + \mathbf{b}_t e_t^*$. $o(k)$
9. Compute: $\gamma_t = \Delta e_t^* |e_t|^2$. $o(1)$
10. Compute: $\alpha_t = \alpha_{t-1} - \gamma_t$. $o(1)$
11. If $t < t_o$
12. Compute: $\mathbf{w}_t = \mathbf{w}_{t-1}^*$.
13. Else if $t = t_o$
14. Compute: $\mathbf{Z}_t = \{\lambda^{-1}\mathbf{T}_{t-1} + \gamma_t / [(\alpha_{t-1} - \gamma_t)\alpha_{t-1}]\mathbf{I}_k\}^{-1}$ $o(k^3)$
15. Else
16. Compute $\mathbf{Z}_t$ using Eq. (34). $o(k^2)$
17. End if
18. Compute: $\mathbf{T}_t^{-1}$ using Eq. (35). $o(k^2)$
19. Compute: $\mathbf{w}_t$ using Eq. (36). $o(k^2)$
20. End if
21. Update $\hat{\mathbf{R}}_{1,t}^{-1} = \hat{\mathbf{R}}_{1,t-1}^{-1} - \mathbf{b}_t \hat{\mathbf{x}}_t^T \hat{\mathbf{R}}_{1,t-1}^{-1}$ $o(k^2)$
22. End for
23. Return $\mathbf{w}_t, \mathbf{T}_t^{-1}, \hat{\mathbf{R}}_{1,t}^{-1}$.

$$+ \mathbf{T}_{t-1}^{-1}\big]^{-1} \lambda \mathbf{T}_{t-1}^{-1}$$
$$\approx \lambda \mathbf{T}_{t-1}^{-1} - \lambda^2 \mathbf{T}_{t-1}^{-1}(\gamma_t/((\alpha_{t-1} - \gamma_t)\alpha_{t-1}))\mathbf{T}_{t-1}^{-1} \quad (34)$$

$$\mathbf{T}_t^{-1} = \mathbf{Z}_t + (\mathbf{Z}_t \mathbf{b}_t \hat{\mathbf{x}}_{1,t}^T (\hat{\mathbf{R}}_{1,t-1}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)\mathbf{Z}_t)/1$$
$$- \hat{\mathbf{x}}_{1,t}^T (\hat{\mathbf{R}}_{1,t-1}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)\mathbf{Z}_t \mathbf{b}_t. \quad (35)$$

The final form of $\mathbf{w}_t$ is obtained as

$$\mathbf{w}_t = \mathbf{w}_t^* - (\hat{\mathbf{R}}_{1,t}^{-1}/\|\hat{\mathbf{w}}_{t-1}\|_2^2)\mathbf{T}_t^{-1}\mathbf{w}_t^*. \quad (36)$$

The parameter γ_t gradually converges to zero through successive iterations. Accordingly, (34) can be simplified, resulting in a PI method with improved computational efficiency and enhanced noise robustness. However, under noisy conditions, the Kw must be carefully selected, as an inappropriate value may cause filter divergence and amplify PI errors. To address this issue, a KwA mechanism is introduced to mitigate errors arising from parameter mismatch. As shown in (14), total error information is produced when noise affects the input and output signals. This information is used both as a criterion for optimizing Kw and as the error term in (12) for dynamic adjustment, achieving an optimization effect analogous to that described in Section II-B. Algorithm 1 summarizes the principal computational steps of the PI and provides a detailed evaluation of the computational complexity at each stage, where $o(\cdot)$ denotes terms of the same infinitesimal order.

Remark 3: According to Algorithm 1, a matrix inversion is required only at time t_o, yielding a computational complexity of $o(k^3)$. In all subsequent iterations, the computational cost per step remains at or below $o(k^2)$.

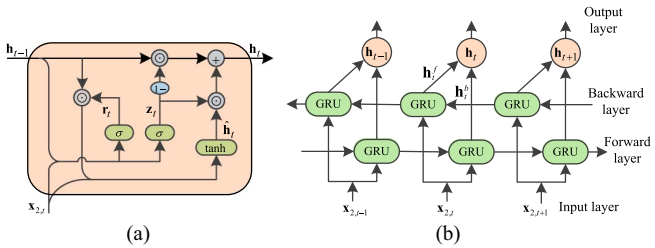

Fig. 2. Structures of (a) GRU model and (b) Bi-GRU model.

III. ST Estimation Based on PINE-BI-GRU Model

A. Overview of ELM Model

The ELM, a single-hidden-layer FNN, is characterized by a distinctive training mechanism. The input-to-hidden layer weights and biases are randomly initialized and remain fixed, while the hidden-to-output weights are analytically determined via the least squares method [13]. This randomization enables efficient mapping of inputs into a HD feature space, facilitating the modeling of complex nonlinear relationships [14]. Its mathematical formulation is given by

$$\begin{cases} \mathbf{H} = \delta(\boldsymbol{\omega} \cdot \mathbf{x} + \mathbf{b}) \\ \mathbf{Y} = \beta \cdot \mathbf{H}. \end{cases} \quad (37)$$

The input, output, and hidden layer output of the ELM are denoted by $\mathbf{x}, \mathbf{Y}$, and $\mathbf{H}$, respectively. The connection weights and biases between the input and hidden layers are represented by $\boldsymbol{\omega}$ and $\mathbf{b}$, while $\delta(\cdot)$ denotes the sigmoid activation function, and β denotes the connection weights between the hidden and output layers. In ELM, the hidden layer weights are randomly initialized and subjected to a nonlinear transformation that maps the input features into a HD space, thereby significantly enhancing data representation. This process eliminates the need for iterative training and substantially improves computational efficiency. Additionally, due to its inherent nonlinearity, ELM effectively captures complex dynamic input/output relationships, enabling robust extraction of nonlinear, HD dynamic features from the input variables.

B. Overview of the Bi-GRU Model

The GRU, an enhanced RNN architecture, improves the processing of time-series data by incorporating update and reset gates, which effectively address gradient vanishing and explosion during training [10], [15]. The update gate adaptively controls the influence of past information on the current state, while the reset gate filters out irrelevant historical data, thereby enhancing the model's ability to capture sequential dependencies. The structure of the GRU is shown in Fig. 2(a), and its mathematical formulation is given as

$$\mathbf{z}_t = \delta(\boldsymbol{\omega}_z \mathbf{x}_{2,t} + \boldsymbol{\varepsilon}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \quad (38)$$
$$\mathbf{r}_t = \delta(\boldsymbol{\omega}_r \mathbf{x}_{2,t} + \boldsymbol{\varepsilon}_r \mathbf{h}_{t-1} + \mathbf{b}_r) \quad (39)$$
$$\hat{\mathbf{h}}_t = \tanh(\boldsymbol{\omega}_h \mathbf{x}_{2,t} + \boldsymbol{\varepsilon}_h (\mathbf{r}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \quad (40)$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \hat{\mathbf{h}}_t. \tag{41}$$

At time t, the current input data is denoted by $\mathbf{x}_{2,t}$, and the current candidate layer state is represented by $\hat{\mathbf{h}}_t$. The hidden states at the current and previous time steps are designated as $\mathbf{h}_t$ and $\mathbf{h}_{t-1}$, respectively. Within the model, the input connection weights for the update gate, reset gate, and candidate layer are defined as $\boldsymbol{\omega}_z$, $\boldsymbol{\omega}_r$, and $\boldsymbol{\omega}_h$, respectively. Similarly, the recurrent weights and biases for the update gate, reset gate, and candidate layer are specified as $\boldsymbol{\varepsilon}_z$, $\mathbf{b}_z$, $\boldsymbol{\varepsilon}_r$, $\mathbf{b}_r$, $\boldsymbol{\varepsilon}_h$, and $\mathbf{b}_h$, respectively. The sigmoid activation function and hyperbolic tangent function are denoted by $\delta(\cdot)$ and $\tanh(\cdot)$, respectively, while the element-wise multiplication operator is represented by $\odot$. The GRU is well-regarded for effectively addressing long-term dependency issues [10], [13], and offers advantages over LSTM networks, including reduced parameter complexity and improved computational efficiency. However, a standard GRU processes sequences in a single temporal direction, limiting its ability to incorporate contextual information from both past and future, which may hinder prediction accuracy. To address this limitation, the Bi-GRU employs a bidirectional architecture consisting of forward and backward GRUs [17], [18]. The forward GRU captures temporal features from past to future, while the backward GRU extracts information in the reverse direction. By concatenating their outputs, a more comprehensive representation of sequential dependencies is obtained. The structure of the Bi-GRU is shown in Fig. 2(b), and its mathematical formulation is expressed as

$$\mathbf{h}_t = \mathbf{h}_t^f \oplus \mathbf{h}_t^b \tag{42}$$

where $\mathbf{h}_t^f$ and $\mathbf{h}_t^b$ are the forward and backward output vectors, respectively, and $\oplus$ represents the vector concatenation operation. The Bi-GRU network is adopted as the core architecture for SE of LIBs due to its inherent ability to extract temporal features and model dynamic systems. By incorporating bidirectional cascaded layers, the proposed framework captures both forward and backward temporal dependencies in multivariate time-series data (e.g., voltage, current, temperature), thereby improving the reconstruction of the nonlinear electrochemical dynamics underlying LIB degradation. The effectiveness of Bi-GRU in SE tasks stems from its parameterized gating mechanisms, which adaptively regulate information flow across time steps, enabling the modeling of both short-term variations and long-term degradation trends in battery behavior.

C. Proposed Method

LIBs, as nonlinear systems, exhibit complex and diverse nonlinear characteristics, with varying correlations between different battery data and ST. Directly using all data for ST estimation enlarges the nonlinear feature space, reducing model identifiability and performance [5], [14]. To mitigate the degradation in identifiability caused by the expanded nonlinear feature space, we propose an enhanced Bi-GRU framework, called PINE-Bi-GRU, integrating physical information and nonlinear extension for high-precision ST estimation. Physical variables, including R_T and SOC, are combined with voltage, current, and AT(T_a) as input features. PCA segments the input into subsequences with significant positive or negative correlations, which are mapped into a HD nonlinear feature space via the hidden layer of an ELM, producing two extended feature arrays. These HD features are fused in parallel with the original LD data to form the Bi-GRU inputs, enabling effective capture of global input–output associations. Collaborative modeling across HD and LD spaces significantly improves ST estimation accuracy. Detailed descriptions of the design and implementation of the proposed framework are provided in the following sections.

1) Step 1: Construction of Positive and Negative Subinput Variables Based on Partial Correlation Analysis: The PCA theory is adopted to distinguish input variables into positive and negative subinputs. The sample set $((\mathbf{x}_{3,i}, \mathbf{y}), \mathbf{x}_3 = [I, T_a, V, R_T, \text{SOC}], i = 1, 2, \ldots, m)$, composed of m-dimensional input variables, includes $\mathbf{x}_{3,i} = (x_{3,i1}, x_{3,i2}, \ldots x_{3,in})$, which corresponds to n samples of the ith input dimension, and $\mathbf{y}$ represents the associated output variable. The correlation between the input and output variables is precisely quantified by

$$r_{xy} = \frac{\sum(\mathbf{x}_{3,i} - \overline{x}_3)(\mathbf{y} - \overline{y})}{\sqrt{\sum(\mathbf{x}_{3,i} - \overline{x}_3)^2(\mathbf{y} - \overline{y})^2}}. \tag{43}$$

The mean values of the input subsequence and the output variable are defined as $\overline{x}_3$ and $\overline{y}$, respectively. Based on the relationship between the input and output variables, a correlation coefficient matrix of dimension $(m + 1)$ is defined

$$\mathbf{r} = (r_{ij})_{(m+1)\times(m+1)} = \begin{bmatrix} r_{11} & \ldots & r_{1m} & r_{1y} \\ \vdots & & \vdots & \vdots \\ r_{m1} & \ldots & r_{mm} & r_{my} \\ r_{y1} & \ldots & r_{ym} & r_{yy} \end{bmatrix} \tag{44}$$

where r_{1m} is the correlation coefficient between variables $\mathbf{x}_{3,1}$ and $\mathbf{x}_{3,m}$ satisfying $r_{1m} = r_{m1}$. The partial correlation coefficients (PCC) between the input and output variables are obtained from the inverse of (44)

$$\mathbf{r}^{-1} = (\alpha_{ij})_{(m+1)\times(m+1)} = \begin{bmatrix} \alpha_{11} & \ldots & \alpha_{1m} & \alpha_{1y} \\ \vdots & & \vdots & \vdots \\ \alpha_{m1} & \ldots & \alpha_{mm} & \alpha_{my} \\ \alpha_{y1} & \ldots & \alpha_{ym} & \alpha_{yy} \end{bmatrix} \tag{45}$$

$$pr_{(x_i,y)} = \frac{-\alpha_{yi}}{\sqrt{\alpha_{ii}\alpha_{yy}}} = \frac{-\alpha_{iy}}{\sqrt{\alpha_{ii}\alpha_{yy}}} \quad (i = 1, 2, \ldots, m) \ . \tag{46}$$

Through an in-depth analysis of the PCC results, the input data will be categorized into two subsets based on differences in correlation. The positively correlated subset $\mathbf{x}_{3,P} = (\mathbf{x}_{3,1}, \mathbf{x}_{3,2}, \ldots, \mathbf{x}_{3,m1})$ is defined by $m1$ variables, whereas the negatively correlated subset $\mathbf{x}_{3,N} = (\mathbf{x}_{3,m1+1}, \mathbf{x}_{3,m1+2}, \ldots, \mathbf{x}_{3,m})$ consists of $m - m1$ variables. Notably, all variables within a given subset exhibit identical partial correlations with the output. As a result, the impact of variables within each subset on the output is considered equivalent.

2) Step 2: The Nonlinear Feature Expansion of Subinput Data is Achieved Through the Use of the ELM Framework: Based on the preceding analysis, the input variables are categorized into

two subsets, which are processed through their respective ELM network architectures to extract nonlinear extended features. The ELM for the positive subset incorporates j_1 hidden layer nodes, while the ELM for the negative subset is configured with j_2 hidden layer nodes. Consequently, weight and bias vectors of dimensions $j_1 \times m1$ and $j_2 \times (m - m1)$ are generated for the positive and negative subsets, respectively. Therefore, the outputs of the hidden layers for these subsets within the ELM framework can be formulated as

$$\begin{cases} \mathbf{H}^P = \delta(\boldsymbol{\omega}^P \cdot \mathbf{x}_{3,P} + \mathbf{b}^P) \\ \mathbf{H}^N = \delta(\boldsymbol{\omega}^N \cdot \mathbf{x}_{3,N} + \mathbf{b}^N) \end{cases} \tag{47}$$

where $\mathbf{H}^P$ and $\mathbf{H}^N$ denote the hidden layer outputs of the positive and negative sub-networks, respectively, while $\boldsymbol{\omega}^P$, $\mathbf{b}^P$, $\boldsymbol{\omega}^N$, and $\mathbf{b}^N$ represent the connection weights and biases from the input layer to the hidden layer for the positive and negative subnetworks, respectively.

3) Step 3: An Enhanced ST Estimation Model via Bi-GRU Incorporating Physical Information and Nonlinear Extensions: By integrating the outputs of the positive and negative ELM hidden layers with the original input data in parallel as the input to the Bi-GRU, regression prediction is achieved. The calculation process is given as

$$\begin{aligned} \mathbf{x}_2 &= [\mathbf{H}^P, \mathbf{H}^N, \mathbf{x}_3]^T \\ &= \begin{bmatrix} h_{11}^P & \cdots & h_{1k_1}^P & h_{11}^N & \cdots & h_{1k_2}^N & x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ h_{n1}^P & \cdots & h_{nk_1}^P & h_{n1}^N & \cdots & h_{nk_2}^N & x_{n1} & \cdots & x_{nm} \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{x}_{2,1} & \cdots & \mathbf{x}_{2,5m} \end{bmatrix}^T \end{aligned} \tag{48}$$

$$\mathbf{h}_t = f(\mathbf{h}_t^f, \mathbf{x}_2) \oplus f(\mathbf{h}_t^b, \mathbf{x}_2) \tag{49}$$

$$\mathbf{y}_t = \mathbf{W}^o \cdot \mathbf{h}_t + \mathbf{b}^o. \tag{50}$$

In the PINE-Bi-GRU model, $\mathbf{x}_2$ and $\mathbf{h}_t$ are defined as the input vector and the output vector, respectively. The computational process of the GRU unit is represented by $f(\cdot)$, where $\mathbf{W}^o$ and $\mathbf{b}^o$ denote the weight and bias of the fully connected layer, respectively, and $\mathbf{y}_t$ indicates the ST estimate.

The computational steps of the PINE-Bi-GRU framework are summarized as follows. First, a diverse battery dataset is constructed by integrating multidimensional data, including equivalent internal resistance derived from physics-based modeling, estimated SOC, and original measurements of voltage, current, and AT. The resulting dataset is employed as input to the Bi-GRU model. In particular, R_T, serving as a physics-guided feature, is incorporated into NN architectures that model temperature-sensitive parameters and dynamic system identifiers reflecting state-dependent over-potentials. Incorporation of this feature improves estimation accuracy across a wide range of thermal conditions while preserving interpretability. Second, partial correlations between input and output variables are analyzed, and the input data are partitioned into positive and negative subinputs. Each subinput is subsequently projected into an HD nonlinear feature space through the hidden layer of an ELM, enabling the extraction of enriched HD representations. Finally, nonlinear extended features from both categories

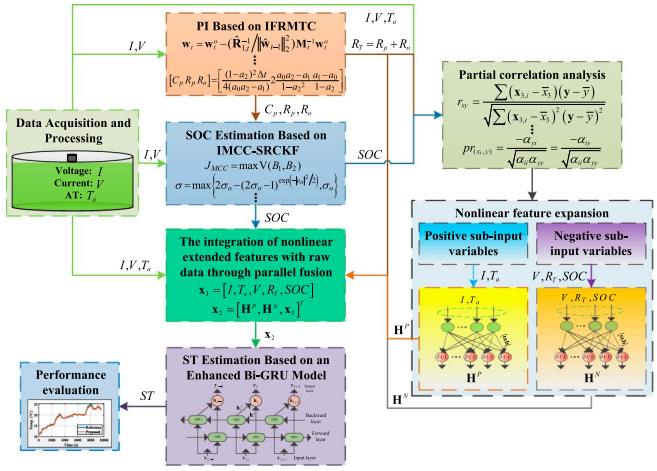

Fig. 3. PINE-Bi-GRU-based robust joint SE framework for LIBs.

of subinputs are fused in parallel with the original input variables and integrated into the Bi-GRU framework for training and testing. The fusion mechanism captures global features spanning HD and LD spaces, considerably strengthening the robustness and generalization capability of the framework and yielding accurate ST estimation. A detailed illustration of the implementation procedure is provided in Fig. 3.

IV. Experiment and Discussion

The proposed estimation methods for PI, SOC, and ST were validated using two publicly available datasets provided by Dr. Phillip Kollmeyer [26], [27], [28]. These datasets were obtained from a Panasonic 18650PF cell (2.9 Ah nominal capacity) and a Samsung INR21700 30T cell (3 Ah nominal capacity). Experiments were performed using a programmable battery tester integrated with a temperature-controlled chamber and a computer-based data acquisition system. Detailed experimental procedures are provided in [26], [27], [28]. According to the dataset documentation, when the ambient temperature fell below 10 °C, the regenerative current was deliberately restricted to prevent premature cell degradation. Consequently, under low-temperature conditions, the current profiles mainly exhibited negative values (discharge direction), while positive current components (charging) were effectively suppressed, a procedure also reported in [9]. Fig. 4 shows a schematic of the experimental platform and its integrated data acquisition system. During testing, voltage, current, and AT were recorded in real time, while ST was continuously monitored using thermocouples. The experiments covered several dynamic driving cycles, including Los Angeles 92 (LA92), highway fuel economy test (HWFET), urban dynamometer driving schedule (UDDS), supplemental federal test procedure driving schedule (US06), and hybrid cycles (Cycle and NN). The dynamic temperature tests spanned ambient conditions from –20 °C to 25 °C for the Panasonic 18650PF cell and from –20 °C to 40 °C for the Samsung INR21700 30T cell. For both cells, OCV-SOC relationships were established from pulse discharge tests conducted at a constant temperature of 25 °C. To enhance data processing efficiency, the raw data were resampled to 1-s

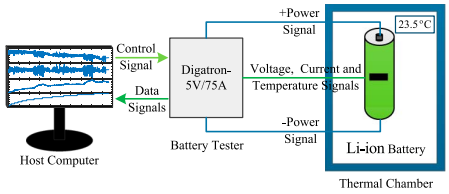

Fig. 4. Schematic representation of the experimental platform and its corresponding data acquisition system.

intervals and normalized to the [0, 1] range for both training and testing datasets. According to findings in [20], [24], under NGN conditions, extended KF, unscented KF, and CKF-based methods exhibit significantly degraded SOC estimation performance, with CKF performing comparatively better. Therefore, the proposed IMCC-SRCKF method is adopted for focused simulation analysis to rigorously evaluate its robustness and accuracy.

Due to the absence of AT data for the Samsung INR21700 30T battery, Mixed 1, Mixed 3 and Mixed 5 at 0 °C and Mixed 2, Mixed 4 and Mixed 6 at 25 °C were used for model training. The input variables included voltage, current, estimated SOC, and R_T, with ST as the output. For the Panasonic 18650PF battery, Cycle 1 and the LA92 driving profile at −20 °C, together with Cycle 3 and the HWFET profile at 0 °C, were employed as training data. Inputs consisted of voltage, current, estimated SOC, R_T, and AT (T_a), while ST served as the output variable. Since the initial SOC and battery model parameters were unknown at the beginning of the experiments, the initial SOC was set to 80%, and the battery parameters were initialized to zero to comprehensively evaluate the overall performance of the proposed PI and SOC estimation methods. The reference SOC in the experiments was calculated using the ampere-hour integration method. To further quantify the performance of the method, a comprehensive analysis was conducted using three metrics: mean absolute error (MAE), root mean square error (RMSE), and maximum error (MAX). All simulation experiments were performed on a computer equipped with an i7-13620H CPU (2.40 GHz) using MATLAB R2022b software

$$\begin{cases} \text{MAE} = \frac{1}{M}\sum_{t=1}^{M}|\phi_t - \phi_t^*| \\ \text{MAX} = \max|\phi_t - \phi_t^*| \\ \text{RMSE} = \sqrt{\frac{1}{M}\sum_{t=1}^{M}(\phi_t - \phi_t^*)^2} \end{cases} \tag{51}$$

where ϕ_t and ϕ_t^* represent the true and the estimated values of the battery state at time t, respectively, with M denoting the number of test samples.

A. Model Validation and SOC Estimation

In this section, an experimental setup distinct from those used in previous studies is established to assess the robustness of the proposed method. To emulate a realistic noisy environment, noise is injected into the original voltage and current signals. Following the methodologies presented in [23], [24], [25], the noise distribution is formulated using expression $v_t = (1-a_t)b_t^* + a_t c_t$, wherein the parameter b_t^* is modeled as a uniformly distributed random variable within the interval [–3

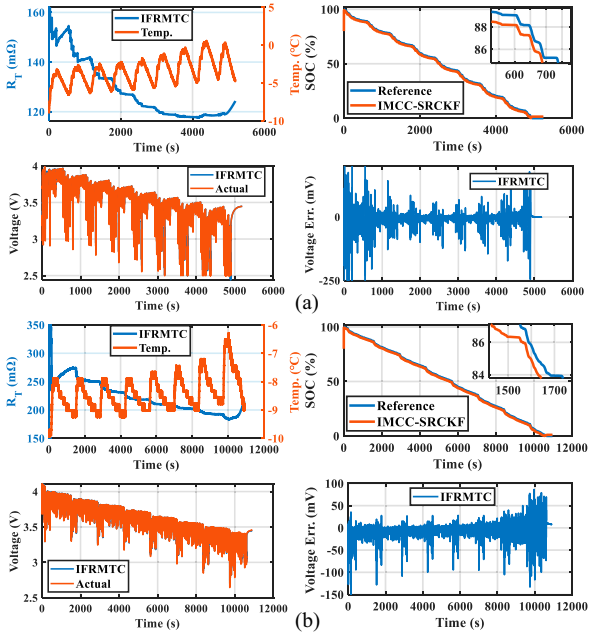

Fig. 5. PI results and model validation under an AT of –10 °C. (a) NN test results. (b) UDDS test results.

3] mV^2/mA^2, and is introduced to represent the background noise component. c_t is specified as GN with a zero mean and a variance of 50 mV^2/mA^2, employed to represent outliers. a_t is designated as an independently and identically distributed binary signal, utilized to control the occurrence probability of outliers, as governed by equation $\Pr(a_t = 0) = 1 - p(0 \le p \le 1)$, $\Pr(a_t = 1) = p$. In this experimental setup, the probability $p = 0.001$.

This section draws on the Panasonic 18650PF battery dataset, and the proposed method is rigorously evaluated under noisy interference conditions. Specifically, model parameters were identified using the IFRMTC algorithm under simultaneous voltage and current noise, upon which a first-order RC-ECM was constructed. The accuracy of PI was validated by comparing the simulated voltage, generated by inputting the raw current signal into the model, with the corresponding measured voltage. Subsequently, the validated parameters were incorporated into a dynamic SOC estimation framework based on the IMCC-SRCKF algorithm. Simulation tests were conducted under both isothermal (10 °C) and dynamic temperature conditions (–10 °C and –20 °C), covering typical driving profiles such as UDDS and NN, with the corresponding results presented in Fig. 5. A significant negative correlation between the R_T and ST parameters was observed, providing theoretical support for the estimation of ST. However, early-stage fluctuations in R_T identification, induced by initial parameter uncertainties, adversely affected the accuracy of voltage prediction and SOC estimation. To address this issue, MCC was employed as the unified cost function for both IFRMTC and IMCC-SRCKF, and a KwA mechanism was introduced. By adopting this approach, initialization errors and noise-induced biases were effectively suppressed, enabling highly accurate dynamic estimation of both PI and SOC. The estimation performance under different

TABLE I
PERFORMANCE EVALUATION OF PI AND SOC ESTIMATION RESULTS

		10 °C		−10 °C		−20 °C		SOC ACT (μs)
		MAE	RMSE	MAE	RMSE	MAE	RMSE	
UDDS	Voltage (mV)	4.87	8.11	7.62	13.79	9.65	22.63	5.59
	SOC(%)	0.75	0.76	0.97	0.99	1.02	1.04	
NN	Voltage (mV)	12.37	23.97	16.42	35.09	21.38	45.04	5.62
	SOC(%)	0.77	0.80	1.08	1.12	1.22	1.30	

conditions is summarized in Table I. As the temperature decreased, battery degradation intensified, leading to increased PI deviations and estimation errors. Despite this, even under complex dynamics and extreme environmental conditions, the proposed method exhibited exceptional modeling and SOC estimation capabilities, with voltage prediction MAE and RMSE maintained within 21 and 45 mV, respectively, and SOC estimation within 1.22% and 1.30%, respectively. The average computation time (ACT) per estimation cycle for SOC was 5.6 microseconds (μs), satisfying real-time application requirements.

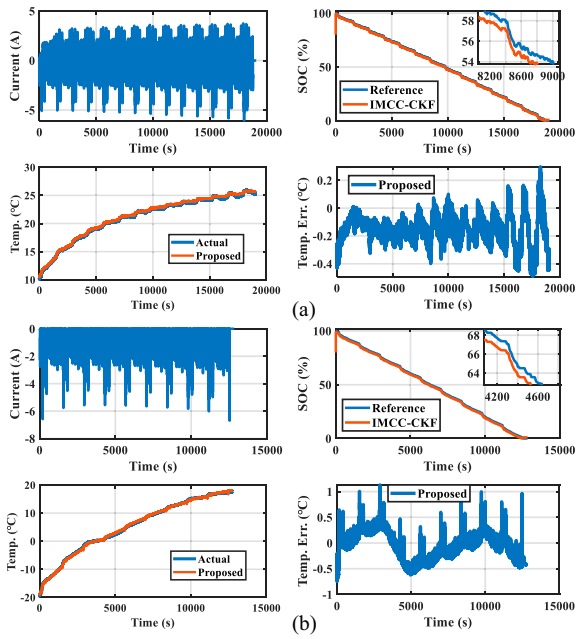

Fig. 6. Test results for the UDDS under varying ATs: current, SOC estimates, ST estimates, and errors. (a) 10°C-UDDS. (b) −20°C-UDDS.

B. Estimation of ST Under Varying ATs

To evaluate the accuracy of the proposed ST estimation method under pronounced and continuous temperature variations, two representative experimental scenarios were designed using the Panasonic 18650PF battery dataset. In these scenarios, the ST was incrementally varied from –20 °C to 20 °C and from 10 °C to 25 °C under controlled temperature conditions. Five input variables, voltage, current, R_T, SOC estimates, and AT (T_a) were systematically evaluated, and their correlations with the target output were analyzed. Through PCC analysis, negative partial correlations of voltage, R_T, and SOC with ST were identified, and these variables were subsequently designated as inputs for the subtractive ELM network. Conversely, a positive partial correlation of current and AT with ST was confirmed, leading to its incorporation into the additive ELM network. To balance computational complexity and prediction accuracy, hyper-parameter settings were optimized via iterative experiments: the number of hidden layer nodes in both subtractive and additive ELM networks was set to 10 for the Bi-GRU main model, single-layer forward and backward GRU hidden layers were configured, each comprising 64 nodes; the Adam optimizer was adopted, with a batch size of 128, a learning rate of 0.001, a gradient threshold of 2, and a maximum of 150 iterations. The efficacy of the proposed ST estimation method was evaluated through a comparative experiment utilizing an improved CNN-LSTM approach [5]. As shown in [5], the accuracy of ICNN-LSTM was significantly enhanced compared with conventional CNN-LSTM and thermal-model-based methods, with negligible variations in computational time, thereby establishing its suitability as a baseline. Additionally, voltage and current data perturbed with Section IV-A noise were introduced to simulate realistic noise conditions. Unlike previous studies, unfiltered ST data were directly employed to evaluate the robustness and generalization capability of the proposed method under noise contaminated scenarios.

Fig. 6 presents the UDDS test results under continuously varying temperature conditions. Despite the presence of complex noise in the voltage and current signals, the SOC estimates were observed to converge rapidly toward the true values and to accurately track the declining SOC trend. Local fluctuations in the ST estimates at specific nodes were identified, attributed to the weakened correlation between temperature and voltage/current signals after prolonged operation, further exacerbated by the highly nonlinear relationship between temperature and internal resistance under environmental influences. To evaluate the generalization capability of the PINE-Bi-GRU model, a comparative analysis with the method reported in [5] was conducted across various driving cycles, and the corresponding numerical results were recorded in Table II. The results indicated that the MAE and RMSE of ST estimation were effectively constrained within 0.5°C, while the MAX remained below 1.5°C. Even under complex operating conditions, both MAE and RMSE consistently remained within 0.75°C. Compared with ICNN-LSTM, the proposed method, by integrating detailed LIB physical information and exploiting nonlinear HD extended features extracted through the ELM hidden layers, effectively captured the global correlations between input and output variables, thereby significantly improving the accuracy of ST estimation. To comprehensively evaluate the differences in computational efficiency and generalization performance between the proposed method and the ICNN-LSTM, comparative experiments were conducted under five distinct operating and temperature conditions. The computation time and the MAE of ST estimation for both methods were recorded, and the corresponding results are shown in Fig. 7. The findings reveal that although the proposed method incurs a slight increase in

TABLE II
PERFORMANCE EVALUATION OF ST ESTIMATION RESULTS UNDER VARYING ATS

	Driving Profiles	PINE-Bi-GRU			ICNN-LSTM [5]		
		RMSE (°C)	MAE (°C)	MAX (°C)	RMSE (°C)	MAE (°C)	MAX (°C)
−20 °C	Cycle 2	0.33	0.31	1.27	0.51	0.35	1.32
	Cycle 4	0.33	0.27	1.31	0.58	0.45	1.63
	NN	0.58	0.48	2.07	0.92	0.82	2.21
	UDDS	0.28	0.23	1.13	0.48	0.41	1.12
	US06	0.75	0.54	2.48	1.96	1.82	3.04
−10 °C	Cycle 2	0.23	0.21	0.45	0.25	0.19	0.72
	Cycle 4	0.18	0.13	0.75	0.25	0.18	0.72
	NN	0.68	0.58	1.47	/	/	/
	UDDS	0.18	0.16	0.49	0.42	0.36	0.98

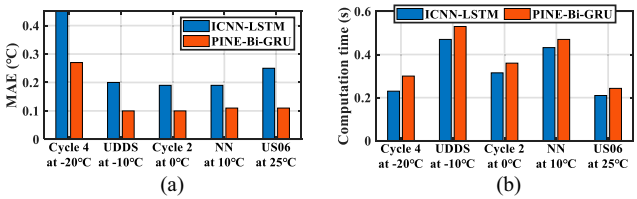

Fig. 7. Comparison of the two estimation methods under varying dynamic and temperature conditions. (a) MAE. (b) Computation time.

computation time under dynamic conditions, its estimation accuracy remains markedly superior to that of the ICNN-LSTM, demonstrating strong generalization capability.

C. Performance Comparison of Different Methods

To evaluate the joint estimation accuracy of PI, SOC, and ST under local temperature transients, the Samsung INR21700 30T battery was selected as the research subject. Validation was carried out under four dynamic operating conditions at two extreme temperatures (40 °C and –20 °C). Dedicated simulation experiments were designed for different estimation objectives. The forgetting factor recursive least squares (FRLS) [3] algorithm was employed as the benchmark for PI estimation, while the multiinnovation SRCKF (MI-SRCKF) [29] was adopted as the reference for SOC estimation. The PI and SOC estimation results derived from the combined FRLS and SRCKF methods were subsequently used as extended inputs to the ST estimation framework to assess the influence of PI and SOC estimation errors on the proposed ST estimation approach, referred to here as Method 1. To further verify the robustness of the proposed approach, an additional set of simulation experiments was conducted. In this case, the raw ST signal was left unfiltered; instead, Gaussian noise with a mean of zero and a variance of 0.2 $°C^2$ was added, and the proposed method was applied for joint SE. This configuration is referred to as Method 2. Meanwhile, voltage and current data corrupted by the noise described in Section IV-A were incorporated to emulate realistic measurement disturbances. Due to the unavailability of AT data, this section focuses solely on examining the latent relationships between ST and the input variables (voltage, current, R_T, and estimated SOC). The PCC analysis revealed that voltage, R_T, and SOC exhibited negative correlations with ST and were therefore assigned as inputs to the negative ELM network. Conversely, current showed a positive correlation with ST and was used as an input to the positive ELM network. Compared with the hyperparameter configuration in Section IV-B, only the learning rate of the Bi-GRU model was adjusted to 0.002, while all other parameters remained unchanged. The results confirmed that, across different temperatures and dynamic operating conditions, the polarity of correlations between the input variables and ST remained stable.

TABLE III
PERFORMANCE EVALUATION OF PI UNDER VARYING TEMPERATURE AND DYNAMIC OPERATING CONDITIONS

		LA92		UDDS		US06		PI
		MAE (mV)	RMSE (mV)	MAE (mV)	RMSE (mV)	MAE (mV)	RMSE (mV)	ACT (μs)
FRLS	−20 °C	25.40	51.41	20.52	40.73	22.15	49.95	0.16
	40 °C	21.86	36.04	18.13	34.30	19.86	42.85	
IFRMTC	−20 °C	7.22	18.94	7.13	17.42	9.02	22.36	0.49
	40 °C	6.54	12.22	6.15	10.18	4.67	13.10	

TABLE IV
PERFORMANCE EVALUATION OF SOC ESTIMATION UNDER VARYING TEMPERATURE AND DYNAMIC OPERATING CONDITIONS

		MI-SRCKF-FRLS		MI-SRCKF-IFRMTC		IMCC-CKF-IFRMTC	
		−20 °C	40 °C	−20 °C	40 °C	−20 °C	40 °C
LA92	MAE (%)	3.55	3.61	2.58	2.52	1.07	0.64
	RMSE (%)	5.06	3.88	4.14	2.93	1.27	0.83
US06	MAE (%)	2.47	1.98	1.55	1.16	1.09	0.76
	RMSE (%)	4.24	2.92	2.75	2.27	1.33	0.93
PI and SOCACT (μs)		4.37	4.34	4.69	4.75	6.19	6.13

Fig. 8(a) and (b) shows the model validation and the SOC and ST estimation results under the LA92 driving cycle at 40 °C and –20 °C, respectively. The corresponding evaluation metrics are summarized in Tables III–V. The results reveal that when both current and voltage are simultaneously influenced by NGN noise, the PI estimation accuracy of the FRLS-based algorithm degrades substantially, leading to significant voltage fluctuations. Under these conditions, the MAE and RMSE of voltage estimation exceed those obtained using the proposed IFRMTC algorithm by factors of 2.4 and 2.2, respectively. When the same SOC estimation algorithm is employed, higher PI accuracy corresponds to smaller SOC estimation errors, in agreement with theoretical expectations. Therefore, precise PI estimation is essential for achieving high-accuracy SE estimation. As illustrated in Fig. 8, the MI-SRCKF algorithm demonstrates limited robustness to NGN interference, preventing rapid convergence to the SOC reference value during the initial stage, even with high-precision PI inputs. In contrast, the proposed method limits the SOC estimation errors to within 1.33% (MAE) and 0.93% (RMSE), while maintaining a PI and SOC average computation time of only 6.19 μs per estimation cycle, thereby meeting real-time performance requirements. For the ST estimation, slight increases in error are observed under local temperature transients and Gaussian noise introduced in Method 2. Nevertheless, the proposed method maintains

TABLE V
PERFORMANCE EVALUATION OF ST ESTIMATION UNDER VARYING TEMPERATURE AND DYNAMIC OPERATING CONDITIONS

	Driving Profiles	Method 2		Method 1		PINE-Bi-GRU	
		MAE (°C)	RMSE (°C)	MAE (°C)	RMSE (°C)	MAE (°C)	RMSE (°C)
−20 °C	HWFET	0.31	0.40	0.20	0.24	0.09	0.11
	LA92	0.27	0.35	0.22	0.24	0.04	0.05
	UDDS	0.29	0.34	0.11	0.12	0.08	0.09
	US06	0.27	0.33	0.14	0.19	0.07	0.09
40 °C	HWFET	0.09	0.11	0.10	0.12	0.02	0.03
	LA92	0.10	0.13	0.05	0.08	0.04	0.05
	UDDS	0.08	0.11	0.09	0.11	0.07	0.07
	US06	0.16	0.21	0.09	0.12	0.04	0.05

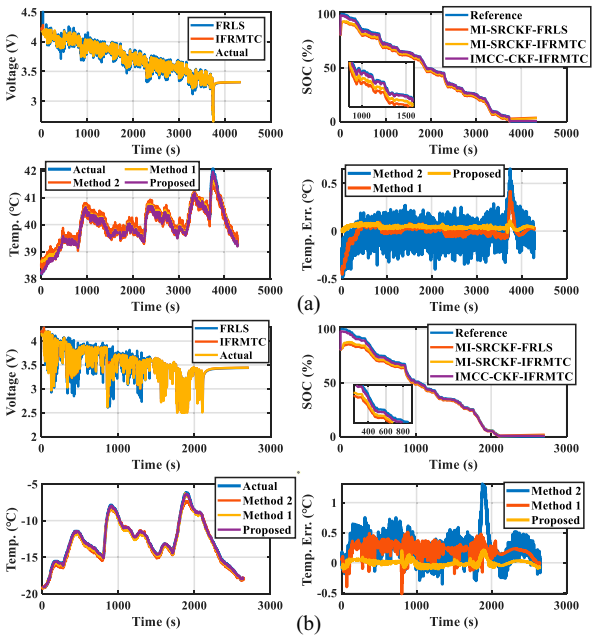

Fig. 8. LA92 Test results at 40 °C and –20 °C: model validation, SOC and ST estimation, and associated errors. (a) −40 °C-LA92. (b) −20 °C-LA92.

stable estimation performance under such challenging conditions, with ST estimation MAE and RMSE remaining below 0.31 °C and 0.4 °C across all four test scenarios. Although Method 1 produces marginally smaller errors, abrupt deviations persist due to localized inaccuracies in PI and SOC estimation. When both PI and SOC estimations achieve maximum precision and no additional noise interference is present, the proposed method attains superior performance. The ST estimation process exhibits minimal fluctuation, with MAE and RMSE restricted to 0.09 °C and 0.11 °C, respectively.

V. CONCLUSION

A robust PI method based on the IFRMTC algorithm was initially proposed. Subsequently, the MCC was employed as the cost function within the SRCKF algorithm, thereby significantly enhancing the accuracy of SOC estimation under noisy conditions. Concurrently, a dynamic KwA adjustment strategy was introduced to enable the online self-adaptation of the Kw parameter, effectively overcoming the challenges associated with its optimization in noise-contaminated environments. Thereafter, physical information was integrated with measurement data to construct a multidimensional feature input set. Subinput structures were derived through PCC, and HD nonlinear features were extracted using the ELM framework, thus substantially augmenting the model's capacity for feature representation. Finally, collaborative modeling of contextual information across HD and LD spaces was realized by simultaneously feeding both the nonlinear extended features and the raw measurement data into the Bi-GRU model. Experimental results demonstrated that, under complex operating conditions, the MAE and RMSE of SOC estimation errors were consistently maintained within 1.33%. Compared with advanced ST estimation methods, the proposed approach achieved average reductions of approximately 50% and 47% in RMSE and MAE, respectively, under scenarios characterized by abrupt temperature fluctuations and complex operational conditions.

In the future, periodic updates of the PINE-Bi-GRU framework will be conducted through capacity and OCV-SOC recalibration, systematic performance evaluations, and comparisons between estimated and measured temperatures so that long-term stability and accuracy can be maintained.

REFERENCES

[1] S. Wang, K. Ou, W. Zhang, and Y. Wang, "A state-of-charge and state-of-health joint estimation method of lithium-ion battery based on temperature-dependent extended Kalman filter and deep learning," *IEEE Trans. Ind. Electron*, vol. 72, no. 1, pp. 570–579, Jan. 2025.

[2] X. Ji, Y. Chen, J. Wang, G. Zhou, C. S. Lai, and Z. Dong, "Time-frequency hybrid neuromorphic computing architecture development for battery state-of-health estimation," *IEEE Internet Things J*, vol. 11, no. 24, pp. 39941–39957, Dec. 2024.

[3] J. Tian et al., "Feature fusion-based inconsistency evaluation for battery pack: improved Gaussian mixture model," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 1, pp. 446–458, Jan. 2023.

[4] X. Zhang, Y. Wang, and Z. Chen, "SOC-modified core temperature estimation of lithium-ion battery based on control-oriented electro-thermal model," *IEEE Trans. Power Electron.*, vol. 38, no. 9, pp. 11642–11651, Sep. 2023.

[5] Y. Zheng, Y. Che, X. Hu, X. Sui, and R. Teodorescu, "Sensorless temperature monitoring of lithium-ion batteries by integrating physics with machine learning," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 2, pp. 2643–2652, Jun. 2024.

[6] 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 Trans. Ind. Electron.*, vol. 68, no. 1, pp. 312–323, Jan. 2021.

[7] X. Cui and B. Xu, "State of charge estimation of lithium-ion battery using robust kernel fuzzy model and multi-innovation UKF algorithm under noise," *IEEE Trans. Ind. Electron.*, vol. 69, no. 11, pp. 11121–11131, Nov. 2022.

[8] Z. Zhao et al., "Online temperature estimation for lithium-ion batteries utilizing a single-frequency impedance unaffected by their peripheral circuits," *IEEE Trans. Power Electron.*, vol. 39, no. 11, pp. 15118–15135, Nov. 2024.

[9] W. Li, Y. Xie, X. Hu, Y. Zhang, H. Li, and X. Lin, "An online SOC-SOTD joint estimation algorithm for pouch Li-ion batteries based on spatio-temporal coupling correction method," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 7370–7386, Jun. 2022.

[10] Q. Yao, D. D. C. Lu, and G. Lei, "A surface temperature estimation method for lithium-ion battery using enhanced GRU-RNN," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 1, pp. 1103–1112, Mar. 2023.

[11] S. Surya, A. Chhetri, and V. Rao, "Kalman filter-machine learning fusion for core temperature estimation in Li-ion batteries," *J. Energy Storage*, vol. 113, Dec. 2025. Art. no. 115656.

[12] Z. Doong, M. Yang, J. Wang, H. Wang, C. S. Lai, and X. Ji, "PFFN: A parallel feature fusion network for remaining useful life early prediction of lithium-ion battery," *IEEE Trans. Transport. Electrific.*, vol. 11, no. 1, pp. 2696–2706, Feb. 2025.

[13] W. Wu and S. Lu, "Remaining useful life prediction of lithium-ion batteries based on data preprocessing and improved ELM," *IEEE Trans. Instrum. Meas.*, vol. 72, Apr. 2023, Art. no. 2510814.

[14] Y. He, P. Wang, and Q. Zhu, "Improved Bi-LSTM with distributed nonlinear extensions and parallel inputs for soft sensing," *IEEE Trans. Ind. Inform.*, vol. 20, no. 3, pp. 3748–3755, Mar. 2024.

[15] L. Shen, J. Li, J. Liu, L. Zhu, and H. Shen, "Temperature adaptive transfer network for cross-domain state-of-charge estimation of Li-ion batteries," *IEEE Trans. Power Electron.*, vol. 38, no. 3, pp. 3857–3869, Mar. 2023.

[16] Z. Dong, X. Ji, J. Wang, Y. Gu, J. Wang, and D. Qi, "ICNCS: internal cascaded neuromorphic computing system for fast electric vehicle state-of-charge estimation," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 4311–4320, Feb. 2024.

[17] L. Chen, Y. Song, A. M. Lopes, X. Bao, Z. Zhang, and Y. Lin, "Joint estimation of state of charge and state of energy of lithium-ion batteries based on optimized bidirectional gated recurrent neural network," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 1, pp. 1605–1616, Mar. 2024.

[18] N. S. Chauhan, N. Kumar, and A. Eskandarian, "A novel confined attention mechanism driven Bi-GRU model for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 9181–9191, Aug. 2024.

[19] W. Zhang, L. Wang, L. Wang, C. Liao, and Y. Zhang, "Joint state-of-charge and state-of-available-power estimation based on the online parameter identification of lithium-ion battery model," *IEEE Trans. Ind. Electron.*, vol. 69, no. 4, pp. 3677–3688, Apr. 2022.

[20] C. Xu, E. Zhang, K. Jiang, and K. Wang, "Dual fuzzy-based adaptive extended Kalman filter for state of charge estimation of liquid metal battery," *Appl. Energy*, vol. 327, Oct. 2022, Art. no. 120091.

[21] Z. Liu, Z. Zhao, Y. Qiu, B. Jing, and C. Yang, "State of charge estimation for Li-ion batteries based on iterative Kalman filter with adaptive maximum correntropy criterion," *J. Power Sources*, vol. 580, May 2023, Art. no. 233282.

[22] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-gaussian signal processing," *IEEE Trans. Signal Process*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[23] F. D. Mandanas and C. L. Kotropoulos, "Robust multidimensional scaling using a maximum correntropy criterion," *IEEE Trans. Signal Process*, vol. 65, no. 4, pp. 919–932, Feb. 2017.

[24] P. Guo, W. Ma, X. Liu, and B. Chen, "Robust SOC estimation of lithium- ion battery combining gated recurrent unit and square root CKF with generalized correntropy," *Energy*, vol. 338, Oct. 2025, Art. no. 138851.

[25] G. Qian, J. Mei, H. H. C. Iu, and S. Wang, "Fixed-point maximum total complex correntropy algorithm for adaptive filter," *IEEE Trans. Signal Process*, vol. 69, pp. 2188–2202, Apr. 2021.

[26] P. Kollmeyer, "Panasonic 18650pf Li-ion battery data," *Mendeley Data*, vol. 1, no. 2018, pp. 1–15, 2018.

[27] E. Chemali, P. Kollmeyer, M. Preindl, R. Ahmed, and A. Emadi, "Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6730–6739, Aug. 2018.

[28] P. Kollmeyer and M. Skells, "Samsung INR21700 30T 3Ah Li-ion battery data," *Mendeley Data*, 2023, doi: 10.17632/9xyvy2njj3.2.

[29] X. Shu, G. Li, Y. Zhang, S. Shen, Z. Chen, and Y. Liu, "Stage of charge estimation of lithium-ion battery packs based on improved cubature Kalman filter with long short-term memory model," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 3, pp. 1271–1284, Sep. 2021.