Hybrid Model Temperature Field Prediction Method Based on Dynamic Mode Decomposition and Deep Learning for IGBT Modules

Jiahao Geng, Fujin Deng, *Senior Member, IEEE*, Qiang Yu, Yaqian Zhang, *Member, IEEE*, Zhe Chen, *Fellow, IEEE*, and Frede Blaabjerg, *Fellow, IEEE*

***Abstract*—The detailed temperature field distribution of high-power insulated gate bipolar transistor (IGBT) modules is important information for the reliability analysis and thermal design of power electronic systems and it is difficult to obtain quickly. This article proposes a hybrid model for temperature field prediction based on dynamic mode decomposition (DMD) and deep learning for IGBT modules. First, the IGBT temperature field (ITF) snapshot is obtained through finite element simulation. Second, the DMD is used to extract stable and unstable trends of the ITF snapshot, and the future stable trend is predicted by recursion. Third, a deep learning model autoencoder-long short-term memory is proposed to predict the future unstable trend. Finally, the ITF prediction snapshots are obtained by adding the future stable and unstable trends. The proposed hybrid model prediction method realizes precise ITF prediction and significantly reduces the computation time. Experimental and simulation results validate the viability of the proposed method.**

***Index Terms*—Dynamic mode decomposition (DMD), finite element method (FEM), insulate-gate bipolar transistor (IGBT), temperature field prediction.**

I. Introduction

INSULATED gate bipolar transistor (IGBT) serves as the fundamental component of power electronic converters due to its simple drive control, high state-current, and low state-voltage [1], [2], [3]. Although IGBTs are highly robust, they can experience catastrophic failures in field applications due to extreme thermal stress [4]. Consequently, precise prediction of the IGBT temperature field (ITF) is crucial for designing inverters that are both safe and efficient [5], [6].

Received 18 October 2024; revised 22 February 2025; accepted 25 March 2025. Date of publication 7 April 2025; date of current version 26 May 2025. This work was supported by the National Natural Science Foundation of China under Grant 52277173 and Grant U23B20137. Recommended for publication by Associate Editor C. DiMarino. *(Corresponding author: Fujin Deng.)*

Jiahao Geng, Qiang Yu, and Yaqian Zhang are with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: gengjiahao@seu.edu.cn; yuqiang@seu.edu.cn; yqzh@seu.edu.cn).

Fujin Deng is with the Jiangsu Key Laboratory of Smart Grid Technology and Equipment, School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: fdeng@seu.edu.cn).

Zhe Chen and Frede Blaabjerg are with the Department of Energy Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: zch@et.aau.dk; fbl@et.aau.dk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2025.3557288.

Digital Object Identifier 10.1109/TPEL.2025.3557288

The methods for ITF prediction can be broadly classified into three categories: measurement method, thermal-impedance model, and numerical method. Measurement methods for the ITF are directly through experiments, including the thermal sensor method and the infrared heat detection method. The thermal sensor method utilizes the characteristic of a thermal element that changes with temperature to measure it, but it must be in complete contact with the semiconductor chip's surface [7]. On the other hand, the infrared heat detection method applies an infrared temperature measuring device to measure the junction temperature of the IGBT module, which does not require contact with the semiconductor chip and can provide temperature distribution across the entire chip surface. However, the thermal sensor method might introduce extra thermal impedance, and the infrared heat detection method requires opening of the IGBT package, which may damage the device or affect its performance [8]. Moreover, this direct measurement method is intrusive and inefficient, and data collection is a challenge [9].

Thermal-impedance models are easy to implement in circuit simulation software, with relatively fast simulation speeds [10], [11], [12], [13], [14], [15]. These models include the Foster and Cauer models [10], normally represented by *RC* networks with thermal resistance and thermal capacitance. The Cauer model is based on the physical structure and material properties of the module, allowing the temperature of each layer to be represented [11]. Therefore, this model is relatively accurate but difficult to construct. In contrast, the Foster model is an equivalent model fitted from measurement or simulation results. The model is easy to establish but can only express the temperature at the measurement points. Additionally, three-dimensional (3-D) *RC* lumped thermal networks have been proposed and improved, considering the thermal coupling effects between the chip and key layers, extending the thermal-impedance model from 1-D to 3-D [12]. However, the 3-D thermal-impedance model is complex due to the need for a series of finite element simulations to fit a large number of parameters [13]. Repeated modeling work is required if the IGBT module or cooling conditions change. Due to these characteristics, thermal-impedance models are mainly used in long-term electro-thermal simulations to obtain junction temperatures for lifetime prediction [12], [15].

Numerical methods have been used in most commercial finite element simulation software, the finite element method (FEM) is the most representative. The FEM discretizes the continuous

region into a finite number of nodes and solve the node temperatures through differential equations to predict ITF. FEM can handle the strong thermal coupling effect of IGBT modules, thus having excellent temperature prediction accuracy [16]. Despite the powerful capabilities of FEM, the heavy computational burden of solving complex differential equations may hinder their applicability [17], [18]. Unfortunately, IGBT modules normally need to be divided into tens of thousands or even millions of nodes to solve ITF through FEM, which requires hours or even tens of hours of computational time [19]. To accelerate FEM, the existing methods can be mainly classified into two categories: 1) invasive methods and 2) noninvasive methods. Invasive methods achieve accelerated computation by significantly reducing the system's degrees of freedom by projecting high-dimensional finite element models onto low-dimensional subspaces. Classical projection techniques, such as the Proper orthogonal decomposition (POD) method [20], Krylov subspace methods [21], and Galerkin projection method [22], are widely employed in the construction of such models. However, these methods require obtaining the full state equations of the FEM simulation and calculating the parameters of all mesh point equations, such as the thermal capacity matrix, thermal stiffness matrix, and thermal load matrix, which brings a significant computational burden to modeling. Noninvasive methods construct reduced-order models through a data-driven method, avoiding the strong reliance on control equations inherent in invasive methods. Classical data-driven methods such as long short-term memory (LSTM) [23], artificial neural networks (ANN) [24], deep neural networks (DNN) [25], and Bayesian estimation (BE) [26] have also been used in combination with the POD method to construct reduced-order models. Although artificial intelligence methods have strong advantages in prediction, capturing long-term evolution characteristics requires a large amount of sample data and extensive training to ensure prediction accuracy.

To address the issues of high modeling complexity, heavy computational burden, and long modeling cycles in existing accelerated FEM methods, this article proposes a hybrid model prediction method based on dynamic mode decomposition (DMD) and deep learning for ITF prediction. The ITF snapshot data from the FEM simulation is used to extract stable and unstable trends of ITF by DMD. Then, the future stable trend is predicted by recursion, and the future unstable trend is predicted by the deep learning model Autoencoder-LSTM (AE-LSTM). Finally, the hybrid model was used to predict future ITF with a satisfactory error level. The primary contributions of the proposed method include: 1) The proposed method significantly shortens the computation time of FEM; 2) The proposed method has higher prediction accuracy compared to the other prediction methods; 3) The proposed method can obtain the entire ITF, not just the junction temperature.

The rest of this article is organized as follows. Section II describes the temperature field of the IGBT module. Section III proposes a hybrid model prediction method based on the DMD and deep learning model for the ITF prediction. Section IV discusses the proposed method. Section V presents experiment studies. Finally, Section VI concludes this article.

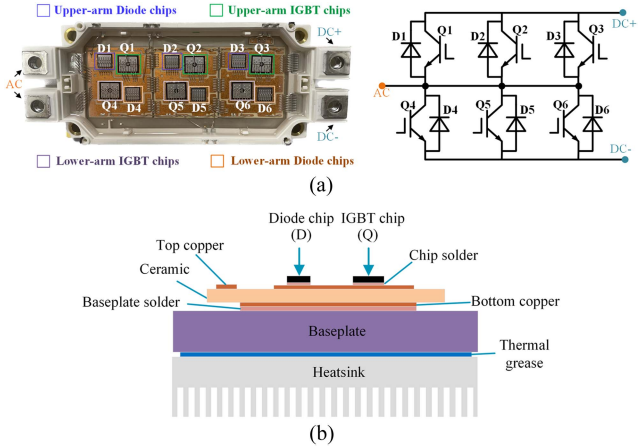

Fig. 1. IGBT module. (a) Physical structure. (b) Principal cross section.

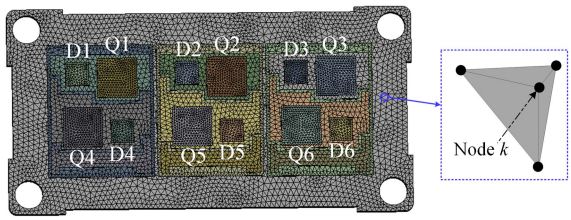

Fig. 2. Finite element model of the IGBT module given in Fig. 1(a).

II. Temperature Field of IGBT Module

This article uses a commercial IGBT module FF300R12ME4 as a sample, rated at 1200 V and 300 A. Fig. 1(a) shows its physical appearance and structural diagram, whereas its principal cross section is shown in Fig. 1(b). The module FF300R12ME4 is a half-bridge configuration, which is composed of an upper arm and a lower arm. The upper arm contains three IGBT chips Q1–Q3 and three diode chips D1–D3 connected in parallel. The lower arm contains three IGBT chips Q4–Q6 and three diode chips D4–D6 connected in parallel. The IGBT chips and diode chips are affixed to a direct bond copper (DBC), which is composed of a top copper layer, a ceramic layer, and a bottom copper layer, using chips solder. This DBC is then attached to a baseplate using baseplate solder. In practical applications, the module is mechanically fixed on the heat sink to enhance heat dissipation efficiency.

To solve the ITF, a FEM discretizes the IGBT module into M nodes, as shown in Fig. 2, numbered as node 1, node 2, …, node M. The temperature T_{n_k} of node k can be solved based on the heat conduction equation [27] as

$$\rho C_p \frac{\partial T_{n\_k}}{\partial t} = \nabla \cdot (K \nabla T_{n\_k}) + \frac{P_{n\_k}}{V_k} \tag{1}$$

where P_{n_k} is the power loss in the region where node k is located, V_k is the volume of the region where node k is located, ∇ is the gradient operator, ρ is the material density, and K is the material thermal conductivity. All node temperatures T_{n_1},

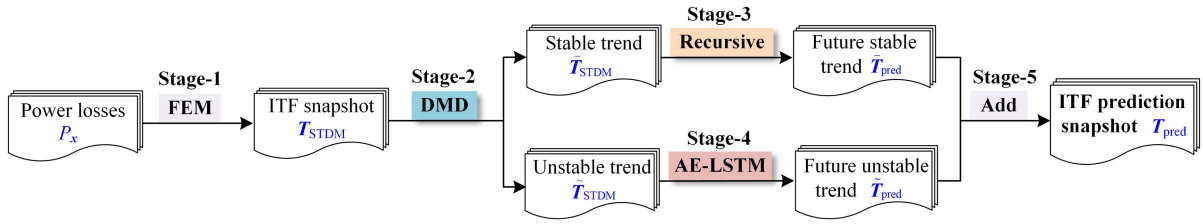

Fig. 3. Proposed hybrid model temperature field prediction method for the IGBT module.

TABLE I
MAJOR MATERIAL OF THE IGBT MODULE

Layers	Materials	ρ (kg/m^3)	K (W/(m·°C))	Cp (J/(kg/°C))
Chip	Si	2330	148	712
Solder	SnAgCu	7430	64	220
DCB copper	Cu	8993	400	385
Ceramic	AL2O3	3800	27	776
Baseplate	Cu	8993	400	385

$T_{n\_2}, \ldots, T_{n\_M}$ constitute the ITF $\boldsymbol{T}_{\mathrm{FEM}}$ as

$$\boldsymbol{T}_{\mathrm{FEM}} = [T_{n\_1}, T_{n\_2}, \ldots, T_{n\_M}]. \tag{2}$$

III. Proposed Hybrid Model Temperature Field Prediction Method

Normally, several hours of computation time are required for the FEM to obtain the ITF, due to the complexity of the FEM model. To reduce the ITF computation time, this article proposes a hybrid model temperature field prediction method using the snapshot data obtained from FEM, as shown in Fig. 3, which includes Stage-1 (ITF snapshot obtaining based on 1FEM), Stage-2 (Stable trend and unstable trend extraction based on DMD), Stage-3 (Future stable trend prediction based on Recursivon), Stage-4 (Future unstable trend prediction based on AE-LSTM). Finally, the future stable and unstable trends are added to form the ITF prediction snapshot.

A. Stage-1: ITF Snapshot Obtained Based on FEM

In this article, the commercial software ANSYS provides a solver for FEM. In ANSYS, the finite element model of the IGBT module shown in Fig. 2 is constructed, and the material parameters are set, as shown in Table I. Normally, the equivalent heat transfer coefficient is used as a simplified thermal boundary condition to represent the heat dissipation capability of the heat sink [28]. The chip (Q1–Q6, D1–D6) regions of the IGBT are set as heat sources, and the power loss P_x of chip x (x = Q1–Q6, D1–D6) is loaded into the heat source.

To verify the accuracy of the FEM model, the junction-case thermal resistance curve from the datasheet is used to compare with the junction-case thermal resistance curve obtained from the FEM model. The junction-case thermal resistance curve Z_{thJC} in the FEM can be calculated by

$$Z_{\mathrm{thJC}}(t) = \frac{T_j(t) - T_c(t)}{P(t)}. \tag{3}$$

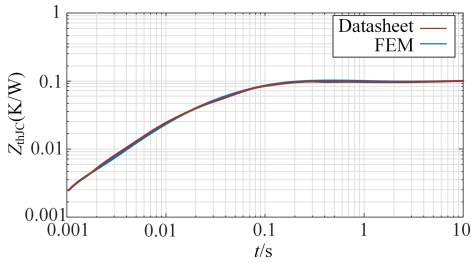

Fig. 4. Thermal resistance curve from the datasheet and the FEM model.

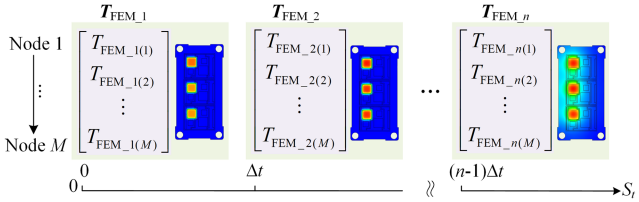

Fig. 5. Sample temperature data matrix $\boldsymbol{T}_{\mathrm{STDM}}$ for ITF.

The junction temperature T_j is the average transient temperature on the surface of the parallel IGBT chips in the FEM simulation model, and the case temperature T_c is the average temperature on the outer surface of the baseplate in the simulation model. Fig. 4 shows the junction-case thermal resistance curves from the datasheet and obtained from the FEM model. It can be seen that the error in the junction-case thermal resistance curves is very low, thus verifying the accuracy of the FEM model.

In ANSYS, the FEM can obtain all node temperatures $\boldsymbol{T}_{\mathrm{FEM}} = [T_{n\_1}, T_{n\_2}, \ldots, T_{n\_M}]$ based on (1) and Table I. Based on $\boldsymbol{T}_{\mathrm{FEM}}$, the IGBT temperature data of nodes 1 to M are sampled with the sampling interval of Δt, as shown in Fig. 5. The temperature data of all nodes at the ith sampling instant is output as a snapshot $\boldsymbol{T}_{\mathrm{FEM}\_i}$, which is expressed as

$$\boldsymbol{T}_{\mathrm{FEM}\_i} = \begin{bmatrix} T_{\mathrm{FEM}\_i(1)} \\ T_{\mathrm{FEM}\_i(2)} \\ \vdots \\ T_{\mathrm{FEM}\_i(M)} \end{bmatrix} \tag{4}$$

where $T_{\mathrm{FEM}\_i(k)}$ is the temperature data of the node k ($1 \le k \le M$) at the ith sampling instant. The n snapshots $\boldsymbol{T}_{\mathrm{FEM}\_1} \sim \boldsymbol{T}_{\mathrm{FEM}\_n}$ are sampled with sampling range $S_t = n \cdot \Delta t$, to form the sample temperature data matrix (STDM) $\boldsymbol{T}_{\mathrm{STDM}}$ as

$$\boldsymbol{T}_{STDM} = [\boldsymbol{T}_{\mathrm{FEM}\_1}, \ldots, \boldsymbol{T}_{\mathrm{FEM}\_i}, \ldots, \boldsymbol{T}_{\mathrm{FEM}\_n}]$$

$$= \begin{bmatrix} T_{\text{FEM\_1}(1)} & T_{\text{FEM\_2}(1)} & \cdots & T_{\text{FEM\_}n(1)} \\ T_{\text{FEM\_1}(2)} & T_{\text{FEM\_2}(2)} & \cdots & T_{\text{FEM\_}n(2)} \\ \vdots & \vdots & \ddots & \vdots \\ T_{\text{FEM\_1}(M)} & T_{\text{FEM\_2}(M)} & \cdots & T_{\text{FEM\_}n(M)} \end{bmatrix}. \tag{5}$$

B. Stage-2: Stable Trend and Unstable Trend Extraction Based on DMD

The STDM $\boldsymbol{T}_{\text{STDM}}$ is used to extract stable trends and unstable trends, which are divided into two time-window matrices $\boldsymbol{T}_{\text{STDM\_}x}$ and $\boldsymbol{T}_{\text{STDM\_}y}$ given as

$$\boldsymbol{T}_{\text{STDM\_}x} = [\boldsymbol{T}_{\text{FEM\_1}}, \ldots, \boldsymbol{T}_{\text{FEM\_}i}, \ldots, \boldsymbol{T}_{\text{FEM\_}n-1}] \tag{6}$$

$$\boldsymbol{T}_{\text{STDM\_}y} = [\boldsymbol{T}_{\text{FEM\_2}}, \ldots, \boldsymbol{T}_{\text{FEM\_}i+1}, \ldots, \boldsymbol{T}_{\text{FEM\_}n}] \tag{7}$$

where $\boldsymbol{T}_{\text{STDM\_}x}$ represents a time window matrix containing snapshots from the first to the nth instant, and $\boldsymbol{T}_{\text{STDM\_}y}$ represents a time window matrix containing snapshots from the second to the $(n+1)$th instant.

In the ITF linear system where material parameter changes are ignored, there exists a best-fit dynamic matrix $\boldsymbol{S}$ that can map approximately $\boldsymbol{T}_{\text{STDM\_}x}$ to $\boldsymbol{T}_{\text{STDM\_}y}$, as

$$\boldsymbol{T}_{\text{STDM\_}y} \approx \boldsymbol{S} \cdot \boldsymbol{T}_{\text{STDM\_}x} \tag{8}$$

with

$$\boldsymbol{T}_{\text{FEM\_}i+1} \approx \boldsymbol{S} \cdot \boldsymbol{T}_{\text{FEM\_}i}. \tag{9}$$

Here, the dynamic matrix $\boldsymbol{S}$ in (9) can be decomposed into superposition of r $(1 \le r \le n)$ order of ITF dynamic modes (DMs) $\varphi_1 \sim \varphi_r$ as

$$\boldsymbol{S} = \sum_{j=1}^{r} \lambda_j \boldsymbol{\varphi}_j \boldsymbol{\varphi}_j^{+} \tag{10}$$

where $\varphi_j{}^+$ is the generalized inverse of φ_j $(j = 1, \ldots, r)$, λ_j is the coefficient of the DM φ_j. The DM φ_j $(j = 1, \ldots, r)$ represents the temperature distribution information over the nodes $1 \sim M$ of the IGBT module at a specific frequency f_j and a specific growth rate g_j, where f_j and g_j can be obtained by

$$\begin{cases} g_j = \text{Re}(\ln(\lambda_j)/\Delta t) \\ f_j = \text{Im}(\ln(\lambda_j)/(\Delta t \times 2\pi)). \end{cases} \tag{11}$$

The DMD is widely used for modes extraction of data in fluid dynamics, disease modeling, and finance, with the advantages of accuracy and efficiency [29]. The DMD is applied to obtain DMs $\varphi_1 \sim \varphi_r$ based on (6)–(10), and the specific calculation is given in the Appendix. In the DMs $\varphi_1 \sim \varphi_r$, as shown in Fig. 6, the DMs with $(g_j = 0)$ and $(f_j = 0)$ are selected as stable DMs φ_s by (11) as

$$\boldsymbol{\varphi}_s = [\boldsymbol{\varphi}_{s1}, \boldsymbol{\varphi}_{s2}, \ldots, \boldsymbol{\varphi}_{sk}] \tag{12}$$

while other modes are classified as unstable modes. The stable modes remain constant and unaffected by time during the evolution of ITF. The unstable modes exhibit oscillatory growth or decay during the evolution of ITF.

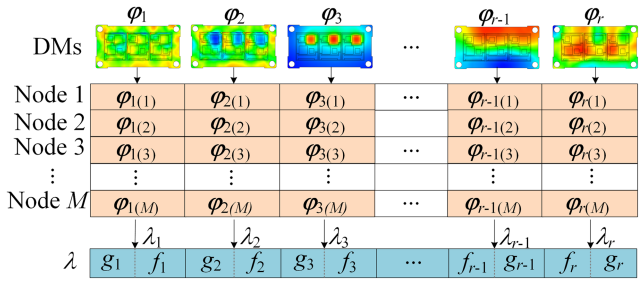

Fig. 6. DMs and the coefficients for the STDM.

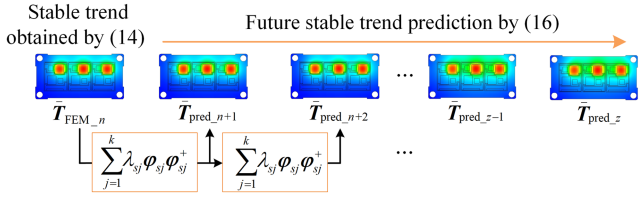

Fig. 7. Prediction principle of the future stable trend.

Based on (9), (10), and (12), the stable trend $\bar{\boldsymbol{T}}_{\text{STDM}}$ and the unstable trend $\tilde{\boldsymbol{T}}_{\text{STDM}}$ of the STDM $\boldsymbol{T}_{\text{STDM}}$ in (5) can be obtained as

$$\begin{cases} \bar{\boldsymbol{T}}_{\text{STDM}} = [\bar{\boldsymbol{T}}_{\text{FEM\_1}}, \ldots, \bar{\boldsymbol{T}}_{\text{FEM\_}i}, \ldots, \bar{\boldsymbol{T}}_{\text{FEM\_}n}] \\ \qquad\qquad = \begin{bmatrix} \bar{T}_{\text{FEM\_1}(1)} & \bar{T}_{\text{FEM\_2}(1)} & \cdots & \bar{T}_{\text{FEM\_}n(1)} \\ \bar{T}_{\text{FEM\_1}(2)} & \bar{T}_{\text{FEM\_2}(2)} & \cdots & \bar{T}_{\text{FEM\_}n(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{T}_{\text{FEM\_1}(M)} & \bar{T}_{\text{FEM\_2}(M)} & \cdots & \bar{T}_{\text{FEM\_}n(M)} \end{bmatrix} \\ \tilde{\boldsymbol{T}}_{\text{STDM}} = \boldsymbol{T}_{\text{STDM}} - \bar{\boldsymbol{T}}_{\text{STDM}} \\ \qquad\qquad = [\tilde{\boldsymbol{T}}_{\text{FEM\_1}}, \ldots, \tilde{\boldsymbol{T}}_{\text{FEM\_}i}, \ldots, \tilde{\boldsymbol{T}}_{\text{FEM\_}n}] \end{cases} \tag{13}$$

with

$$\bar{\boldsymbol{T}}_{\text{FEM\_}i} = \sum_{j=1}^{k} \lambda_{sj}^{(i-1)} \boldsymbol{\varphi}_{sj} \boldsymbol{\varphi}_{sj}^{+} \bar{\boldsymbol{T}}_{\text{FEM\_1}}, (1 \le i \le n) \tag{14}$$

where $\bar{\boldsymbol{T}}_{\text{FEM\_}i}$ and $\tilde{\boldsymbol{T}}_{\text{FEM\_}i}$ is the stable trend snapshot and the unstable trend snapshot at the ith sampling instant, respectively. $\bar{T}_{\text{FEM\_}i(k)}$ is the temperature data of the node k at the ith sampling instant for the stable trend.

C. Stage-3: Future Stable Trend Prediction Based on Recursion

The future stable trend $\bar{\boldsymbol{T}}_{\text{pred}}$ can be predicted by recursion due to its efficient computational performance, as shown in Fig. 7, where the $\bar{\boldsymbol{T}}_{\text{pred}}$ can be obtained based on the stable trend $\bar{\boldsymbol{T}}_{\text{FEM\_}n}$, (9), (10), and (12), as

$$\bar{\boldsymbol{T}}_{\text{pred}} = [\bar{\boldsymbol{T}}_{\text{pred\_}n+1}, \ldots, \bar{\boldsymbol{T}}_{\text{pred\_}i}, \ldots, \bar{\boldsymbol{T}}_{\text{pred\_}z}]$$

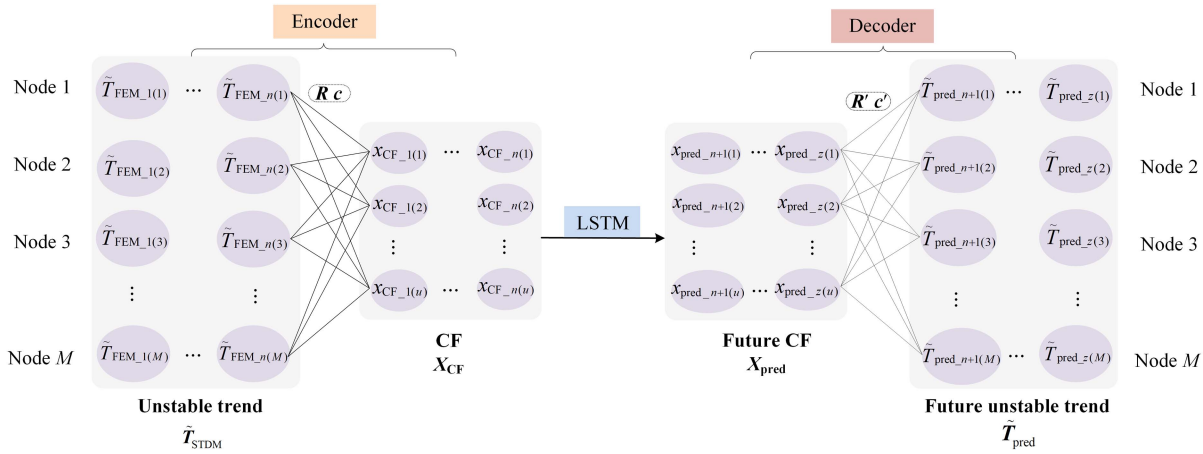

Fig. 8. Proposed unstable trend prediction method based on the AE-LSTM model.

$$= \begin{bmatrix} \bar{T}_{\text{pred\_}n+1(1)} & \bar{T}_{\text{pred\_}n+2(1)} & \cdots & \bar{T}_{\text{pred\_}z(1)} \\ \bar{T}_{\text{pred\_}n+1(2)} & \bar{T}_{\text{pred\_}n+2(2)} & \cdots & \bar{T}_{\text{pred\_}z(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{T}_{\text{pred\_}n+1(M)} & \bar{T}_{\text{pred\_}n+2(M)} & \cdots & \bar{T}_{\text{pred\_}z(M)} \end{bmatrix} \tag{15}$$

with

$$\bar{\boldsymbol{T}}_{\text{pred\_}i} = \sum_{j=1}^{k} \lambda_{sj}^{(i-n)} \varphi_{sj} \varphi_{sj}{}^{+} \bar{\boldsymbol{T}}_{\text{FEM\_}n}, (n+1 \le i \le z) \tag{16}$$

where $\bar{\boldsymbol{T}}_{\text{pred\_}i}$ is the future stable trend snapshot at the ith prediction instant, and $\bar{T}_{\text{pred\_}i(k)}$ is the temperature data of the node k at the ith prediction instant for the future stable trend.

D. Stage-4: Future Unstable Trend Prediction Based on AE-LSTM

The future unstable trend $\tilde{\boldsymbol{T}}_{\text{pred}}$ is proposed to be predicted by the deep learning model AE-LSTM due to its excellent feature extraction and feature learning capabilities [30]. Fig. 8 shows the proposed future unstable trend prediction method based on the AE-LSTM model, which consists of three parts: the encoder of the AE model, the LSTM model, and the decoder of the AE model. In the AE-LSTM model, the compression features (CFs) $\boldsymbol{X}_{\text{CF}}$ for the unstable trend $\tilde{\boldsymbol{T}}_{\text{STDM}}$ obtained from (12) are extracted by encoder first; the future CFs $\boldsymbol{X}_{\text{pred}}$ are predicted by LSTM second; and the future unstable trends $\tilde{\boldsymbol{T}}_{\text{pred}}$ are reconstructed based on the $\boldsymbol{X}_{\text{pred}}$ by Decoder finally, as follows.

1) CFs $\mathbf{X}_{CF}$ Extraction by Encoder: To extract accurate CFs $\boldsymbol{X}_{\text{CF}}$, the weight matrix ($\boldsymbol{R}$, $\boldsymbol{R}$') and bias matrix ($\boldsymbol{c}$, $\boldsymbol{c}$') in the AE model are iteratively optimized by the adaptive learning method [30], until the mean absolute error (MAE) between the new unstable trend $\tilde{\boldsymbol{T}}_{\text{AE}}$ obtained based on the CFs $\boldsymbol{X}_{\text{CF}}$ and the original unstable trend $\tilde{\boldsymbol{T}}_{\text{STDM}}$ is less than given threshold value $\beta_{(\text{AE})}$, as

$$\text{MAE}_{(\text{AE})} = \frac{1}{n} \sum_{i=1}^{n} |\tilde{\boldsymbol{T}}_{\text{FEM\_}i} - \tilde{\boldsymbol{T}}_{\text{AE\_}i}| < \beta_{(\text{AE})} \tag{17}$$

with

$$\begin{aligned} \tilde{\boldsymbol{T}}_{\text{AE}} &= \delta(\boldsymbol{R}' \cdot \boldsymbol{X}_{\text{CF}} + \boldsymbol{c}') \\ &= [\tilde{\boldsymbol{T}}_{\text{AE\_}1}, \ldots, \tilde{\boldsymbol{T}}_{\text{AE\_}i}, \ldots, \tilde{\boldsymbol{T}}_{\text{AE\_}n}] \end{aligned} \tag{18}$$

where $\tilde{\boldsymbol{T}}_{\text{AE\_}i}$ is the new unstable trends at the ith sampling instant. δ is an activation function.

Based on the optimized $\boldsymbol{R}$ and $\boldsymbol{c}$ obtained from (17) and (18), the CFs $\boldsymbol{X}_{\text{CF}}$ for unstable trends $\tilde{\boldsymbol{T}}_{\text{STDM}}$ with dimensions of $u \times n$ can be extracted as

$$\begin{aligned} \boldsymbol{X}_{\text{CF}} &= \delta(\boldsymbol{R} \cdot \tilde{\boldsymbol{T}}_{\text{STDM}} + \boldsymbol{c}) \\ &= [\boldsymbol{X}_{\text{CF\_}1}, \boldsymbol{X}_{\text{CF\_}2}, \ldots, \boldsymbol{X}_{\text{CF\_}n}] \\ &= \begin{bmatrix} x_{\text{CF\_}1(1)} & x_{\text{CF\_}2(1)} & \cdots & x_{\text{CF\_}n(1)} \\ x_{\text{CF\_}1(2)} & x_{\text{CF\_}2(2)} & \cdots & x_{\text{CF\_}n(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\text{CF\_}1(u)} & x_{\text{CF\_}2(u)} & \cdots & x_{\text{CF\_}n(u)} \end{bmatrix} \end{aligned} \tag{19}$$

where $\boldsymbol{X}_{\text{CF\_}i}$ is the CFs at the ith sampling instant, and $x_{\text{CF\_}i(k)}$ is the kth CF at the ith sampling instant.

2) Future CFs $\mathbf{X}_{pred}$ Prediction by LSTM: Fig. 9 shows the future CFs prediction process based on the LSTM model. To accurately predict future CFs $\boldsymbol{X}_{\text{pred}}$, the CFs $\boldsymbol{X}_{\text{CF}}$ in (19) are used as the training data to optimize the LSTM model iteratively. In Fig. 9(a), during the training stage, the CFs $\boldsymbol{X}_{\text{CF}}$ are firstly rearranged using the sliding window (SW) algorithm to generate (n-m) SWs ($\text{SW}_1, \ldots, \text{SW}_i, \ldots, \text{SW}_{n-m}$) of training CFs, where the SW_i consists of the CFs [$\boldsymbol{X}_{\text{CF\_}i}$, $\boldsymbol{X}_{\text{CF\_}i+1}, \ldots, \boldsymbol{X}_{\text{CF\_}i+m\text{-}1}$] from the ith to the (i+m-1)th sampling instant.

In Fig. 9(b), the $\text{SW}_i = [\boldsymbol{X}_{\text{CF\_}i}, \boldsymbol{X}_{\text{CF\_}i+1}, \ldots, \boldsymbol{X}_{\text{CF\_}i+m\text{-}1}]$ are fed into the LSTM model, which consists of LSTM units $i \sim (i+m\text{-}1)$ as shown in Fig. 9(c) and a fully connected layer, and the predicted CF $\boldsymbol{X}_{\text{pred\_}i+m}$ ($i \le n$-m) can be obtained by

$$\boldsymbol{X}_{\text{pred\_}i+m} = \boldsymbol{W}_y \cdot \boldsymbol{h}_{i+m-1} + \boldsymbol{b}_y \tag{20}$$

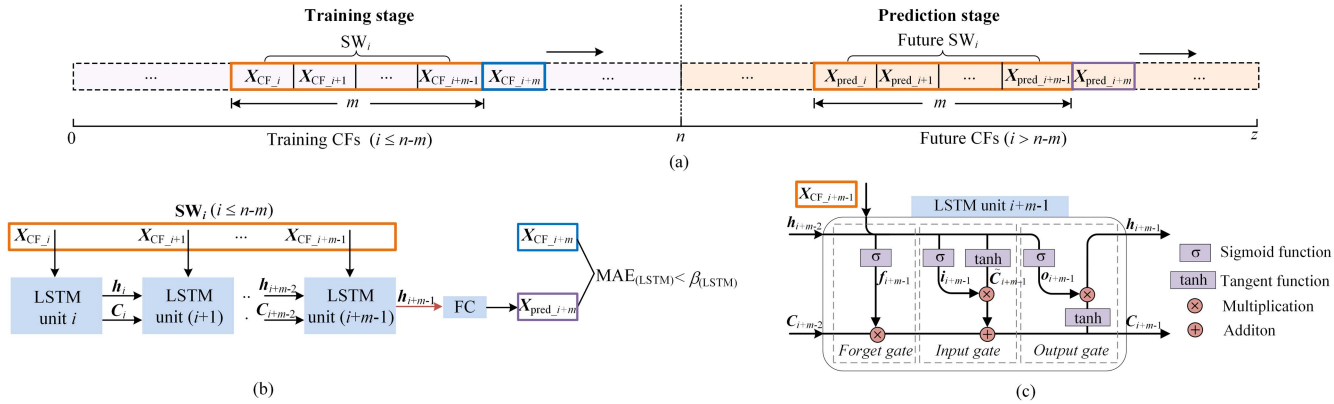

Fig. 9. Future CFs prediction process based on the LSTM model. (a) Sliding window algorithm principle. (b) LSTM model. (c) LSTM unit.

with

$$\begin{cases} \boldsymbol{f}_{i+m-1} = \sigma(\boldsymbol{W}_{fh}\boldsymbol{h}_{i+m-2} + \boldsymbol{W}_{fx}\boldsymbol{X}_{\text{CF\_}i+m-1} + \boldsymbol{b}_f) \\ \boldsymbol{i}_{i+m-1} = \sigma(\boldsymbol{W}_{ih}\boldsymbol{h}_{i+m-2} + \boldsymbol{W}_{ix}\boldsymbol{X}_{\text{CF\_}i+m-1} + \boldsymbol{b}_i) \\ \tilde{\boldsymbol{C}}_{i+m-1} = \tanh(\boldsymbol{W}_{ch}\boldsymbol{h}_{i+m-2} + \boldsymbol{W}_{cx}\boldsymbol{X}_{\text{CF\_}i+m-1} + \boldsymbol{b}_c) \\ \boldsymbol{C}_{i+m-1} = \boldsymbol{f}_{i+m-1} \cdot \tilde{\boldsymbol{C}}_{i+m-2} + \boldsymbol{i}_{i+m-1} \cdot \tilde{\boldsymbol{C}}_{i+m-1} \\ \boldsymbol{o}_{i+m-1} = \sigma(\boldsymbol{W}_{oh}\boldsymbol{h}_{i+m-2} + \boldsymbol{W}_{ox}\boldsymbol{X}_{\text{CF\_}i+m-1} + \boldsymbol{b}_o) \\ \boldsymbol{h}_{i+m-1} = \boldsymbol{o}_{i+m-1} \cdot \tanh(\tilde{\boldsymbol{C}}_{i+m-1}) \end{cases} \tag{21}$$

where the hidden state $\boldsymbol{h}_{i+m\text{-}1}$ and the cell state $\boldsymbol{C}_{i+m\text{-}1}$ store short-term memory information and long-term memory information, respectively. $\boldsymbol{f}_{i+m\text{-}1}$ is the Forget gate filtering out irrelevant information. $\boldsymbol{i}_{i+m\text{-}1}$ is the Input gate decided, which new information is stored. $\boldsymbol{o}_{i+m\text{-}1}$ is the Output gate decided the output information. σ is the sigmoid function. Tanh is the hyperbolic tangent function. The weight matrix $\boldsymbol{W}$ [$\boldsymbol{W}_{fh}$, $\boldsymbol{W}_{fx}$, $\boldsymbol{W}_{ih}$, $\boldsymbol{W}_{ix}$, $\boldsymbol{W}_{ch}$, $\boldsymbol{W}_{cx}$, $\boldsymbol{W}_{oh}$, $\boldsymbol{W}_{ox}$, $\boldsymbol{W}_y$] and the bias matrix $\boldsymbol{b}$ [$\boldsymbol{b}_f$, $\boldsymbol{b}_i$, $\boldsymbol{b}_c$, $\boldsymbol{b}_o$, $\boldsymbol{b}_y$] in (20) and (21) can be iteratively optimized using the adaptive learning rate method, until the $\text{MAE}_{(\text{LSTM})}$ between the predicted CF $\boldsymbol{X}_{\text{pred\_}i+m}$ and the actual CF $\boldsymbol{X}_{\text{CF\_}i+m}$ ($i \le n\text{-}m$) is less than the given threshold value $\beta_{(\text{LSTM})}$, as shown in (22) and Fig. 9(b)

$$\text{MAE}_{(\text{LSTM})} = \frac{1}{n-m}\sum_{i=1}^{n-m} |\boldsymbol{X}_{\text{CF\_}i+m} - \boldsymbol{X}_{\text{pred\_}i+m}| < \beta(\text{LSTM}). \tag{22}$$

In Fig. 9(a), during the prediction stage, based on the optimized $\boldsymbol{W}$ and $\boldsymbol{b}$ obtained from (20)–(22), the future CF $\boldsymbol{X}_{\text{pred\_}i+m}$ ($i > n\text{-}m$) at the ($i+m$)th prediction instant can be obtained by feeding the future $\text{SW}_i = [\boldsymbol{X}_{\text{pred\_}i}, \boldsymbol{X}_{\text{pred\_}i+1}, \ldots, \boldsymbol{X}_{\text{pred\_}i+m\text{-}1}]$ into LSTM model. Finally, the future CFs $\boldsymbol{X}_{\text{pred}}$ ($\boldsymbol{X}_{\text{pred\_}n+1}, \ldots, \boldsymbol{X}_{\text{pred\_}z}$) can be predicted by constantly updating the inputs of the LSTM model using the SW algorithm shown in Fig. 9(a), given as

$$\boldsymbol{X}_{\text{pred}} = [\boldsymbol{X}_{\text{pred\_}n+1}, \ldots, \boldsymbol{X}_{\text{pred\_}i}, \ldots, \boldsymbol{X}_{\text{pred\_}z}] = \begin{bmatrix} x_{\text{pred\_}n+1(1)} & x_{\text{pred\_}n+2(1)} & \cdots & x_{\text{pred\_}z(1)} \\ x_{\text{pred\_}n+1(2)} & x_{\text{pred\_}n+2(2)} & \cdots & x_{\text{pred\_}z(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{\text{pred\_}n+1(u)} & x_{\text{pred\_}n+2(u)} & \cdots & x_{\text{pred\_}z(u)} \end{bmatrix} \tag{23}$$

where the $x_{\text{pred\_}}i(k)$ is the kth future CF at the ith prediction instant.

3) Future Unstable Trends $\tilde{\boldsymbol{T}}_{pred}$ Reconstruction by Decoder: In Fig. 8, based on the future CFs $\boldsymbol{X}_{\text{pred}}$ ($\boldsymbol{X}_{\text{pred\_}n+1}, \ldots, \boldsymbol{X}_{\text{pred\_}z}$) obtained from the LSTM model, the future unstable trends $\tilde{\boldsymbol{T}}_{\text{pred}}$($\tilde{\boldsymbol{T}}_{\text{pred\_}n+1}, \ldots, \tilde{\boldsymbol{T}}_{\text{pred\_}z}$) can be reconstructed by

$$\tilde{\boldsymbol{T}}_{\text{pred}} = \delta(\boldsymbol{R}' \cdot \boldsymbol{X}_{\text{pred}} + \boldsymbol{c}') = [\tilde{\boldsymbol{T}}_{\text{pred\_}n+1}, \ldots, \tilde{\boldsymbol{T}}_{\text{pred\_}i}, \ldots, \tilde{\boldsymbol{T}}_{\text{pred\_}z}] = \begin{bmatrix} \tilde{T}_{\text{pred\_}n+1(1)} & \tilde{T}_{\text{pred\_}n+2(1)} & \cdots & \tilde{T}_{\text{pred\_}z(1)} \\ \tilde{T}_{\text{pred\_}n+1(2)} & \tilde{T}_{\text{pred\_}n+2(2)} & \cdots & \tilde{T}_{\text{pred\_}z(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{T}_{\text{pred\_}n+1(M)} & \tilde{T}_{\text{pred\_}n+2(M)} & \cdots & \tilde{T}_{\text{pred\_}z(M)} \end{bmatrix} \tag{24}$$

where $\tilde{\boldsymbol{T}}_{\text{pred\_}i}$ is the future unstable trend snapshot at the ith prediction instant, and $\tilde{T}_{\text{pred\_}i(k)}$ is the temperature data of the node k at the ith prediction instant for the future unstable trend.

The weight matrix $\boldsymbol{R}'$ and bias matrix $\boldsymbol{c}'$ are obtained by iterating from (17) and (18).

E. Proposed Hybrid Model Prediction Method

Fig. 10 shows the proposed hybrid model for ITF prediction, where the STDM $\boldsymbol{T}_{\text{STDM}}$ is obtained from the FEM simulation. Based on the $\boldsymbol{T}_{\text{STDM}}$ preprocessed by (5) and (6), the DMs φ can be obtained by (28)–(33). Based on the DMs φ, the stable mode φ_s can be selected by (11) and (12). Based on the stable mode φ_s, the stable trend $\bar{\boldsymbol{T}}_{\text{STDM}}$ can be obtained by (14), and the unstable trend $\tilde{\boldsymbol{T}}_{\text{STDM}}$ can be obtained by (13). Afterwards, based on the $\bar{\boldsymbol{T}}_{\text{STDM}}$, the future stable trend $\bar{\boldsymbol{T}}_{\text{pred}}$ can be predicted by (16). On the other hand, the $\tilde{\boldsymbol{T}}_{\text{STDM}}$ is used to train the AE model by

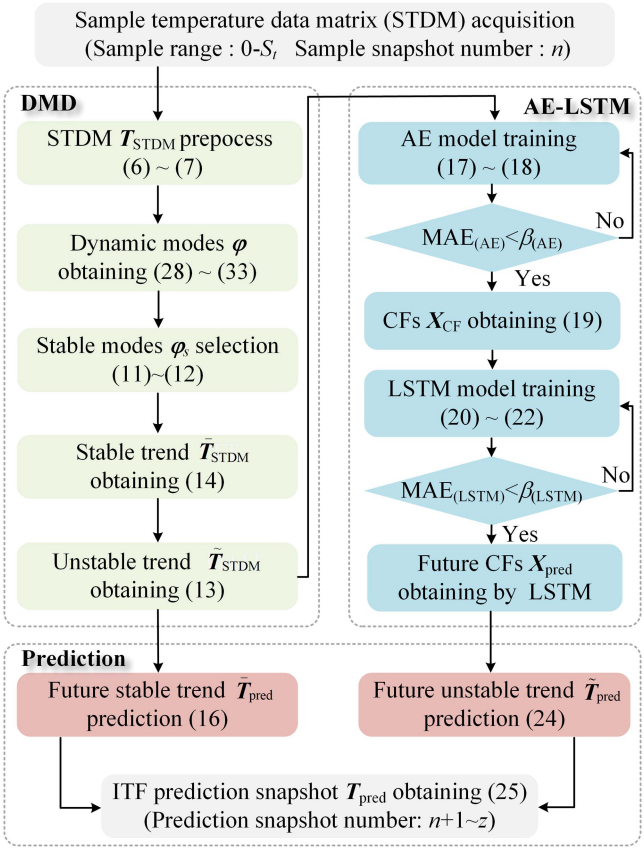

Fig. 10. Flow chart of the proposed method for ITF prediction.

TABLE II
PARAMETERS OF SIMULATION SYSTEM

Parameters	Values
Power loss amplitude P_{peak}	600/W
Equivalent frequency f_p	50/Hz
FEM simulation time step	0.001/s
Sampling interval Δt	0.001/s
Sampling range S_t	3/s
Prediction range t	15/s
Dimensions of the CFs u	20
β_{AE} and β_{LSTM}	$1e^{-12}$ and $1e^{-6}$

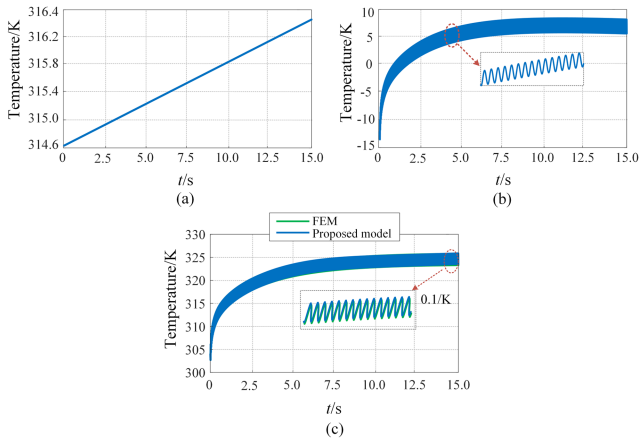

Fig. 11. Prediction results. (a) Stable trend prediction. (b) Unstable trend prediction. (c) Comparison of the results of the proposed method and the FEM.

(17) and (18) until the $\mathrm{MAE}_{(\mathrm{AE})}$ is less than $\beta_{(\mathrm{AE})}$. Based on the trained AE model, the CFs $\boldsymbol{X}_{\mathrm{CF}}$ can be obtained by (19). The CFs $\boldsymbol{X}_{\mathrm{CF}}$ are rearranged by the SW algorithm. Then, they are used to train the LSTM model by (20)–(22) until the $\mathrm{MAE}_{(\mathrm{LSTM})}$ is less than $\beta_{(\mathrm{LSTM})}$. Based on the trained LSTM model and SW algorithm, the future CFs $\boldsymbol{X}_{\mathrm{pred}}$ can be predicted. Afterwards, based on the future CFs $\boldsymbol{X}_{\mathrm{pred}}$, the future unstable trend $\tilde{\boldsymbol{T}}_{\mathrm{pred}}$ can be obtained by (24). Finally, the ITF prediction snapshots $\boldsymbol{T}_{\mathrm{pred}}$ can be obtained by adding the future stable trend $\bar{\boldsymbol{T}}_{\mathrm{pred}}$ in (15) and the future unstable trend $\tilde{\boldsymbol{T}}_{\mathrm{pred}}$ in (24) as

$$\begin{aligned}\boldsymbol{T}_{\mathrm{pred}} &= [\boldsymbol{T}_{\mathrm{pred}\_n+1}, \boldsymbol{T}_{\mathrm{pred}\_n+2}, \ldots, \boldsymbol{T}_{\mathrm{pred}\_z}] \\ &= \bar{\boldsymbol{T}}_{\mathrm{pred}} + \tilde{\boldsymbol{T}}_{\mathrm{pred}}\end{aligned} \tag{25}$$

where $\boldsymbol{T}_{\mathrm{pred}\_i}$ is the ITF snapshot at the ith prediction instant.

IV. Discussion of Proposed Method

A. Performance Evaluation of Proposed Method

In this section, a power loss case is used to evaluate the performance of the proposed method. Normally, power cycling temperature tests are conducted with a single switch consisting of chips connected in parallel [31]. Here, the upper-arm chips consisting of Q1–Q3 in the IGBT module FF300R12ME4 shown in Fig. 1 are used as an example for ITF prediction. To simplify the actual transient power loss of the IGBT module in the inverter, an equivalent power loss curve can be considered [32], which is a sinusoidal half wave as

$$P_{UI} = \begin{cases} P_{\mathrm{peak}} \cdot \sin(2\pi f_p t) & P_{UI} > 0 \\ 0 & P_{UI} \le 0 \end{cases} \tag{26}$$

where P_{UI} is the total power loss of the upper-arm IGBT chips Q1–Q3, P_{peak} is the amplitude, f_p is the equivalent frequency. To simplify the analysis, the power losses between parallel chips Q1–Q3 are considered to be evenly distributed. Thus, the power loss P_x of each chip in the upper-arm IGBT chips is $P_{UI}/3$. Table II shows the used simulation parameters.

The parameters based on Table II were used to predict ITF snapshots by the proposed method and the FEM, respectively. Fig. 11 shows the temperature prediction results at the center point of chip Q2 by the proposed method and the FEM. Fig. 11(a) shows the prediction result of the future stable trend by the DMD. Fig. 11(b) shows the prediction result of the future unstable trend by the AE-LSTM. It can be seen that the future stable trend is a linear rising process, while the future unstable trend is an oscillating process that both rises and falls. Fig. 11(c) shows the prediction results by the proposed method and the FEM. It can be seen that the prediction results of the proposed method and the FEM method are very close to each other.

Fig. 12(a) and (b) shows the ITF snapshot at $t = 15$ s from the proposed method and the FEM, respectively. It can be seen that the temperature responses of the three parallel chips (Q1–Q3) are different due to thermal coupling effects. The highest temperature of chip Q2 is higher than that of chips Q1 and

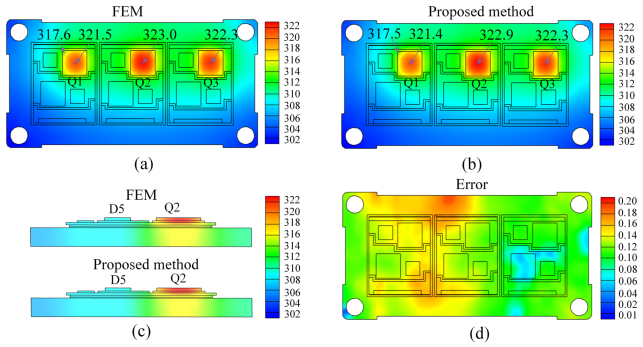

Fig. 12. ITF prediction snapshots. (a) From FEM. (b) From the proposed method. (c) Cross-section temperature snapshot. (d) Error snapshot.

TABLE III
TIME COST OF THE PROPOSED METHOD AND THE FEM METHOD

Method	Stage	Time(s)
Proposed method	Sample snapshots obtaining	3240
	Snapshot trend extraction	0.6
	Stable trend prediction	0.01
	AE-LSTM model training	221
	Unstable trend prediction	2.4
FEM	Simulation	16200

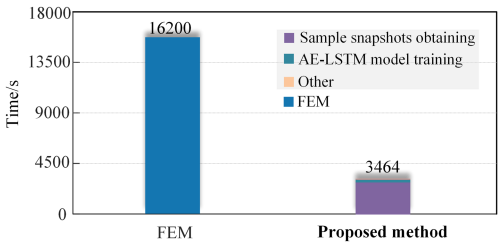

Fig. 13. Time cost of the proposed method and the FEM.

Q3. Additionally, within the same chip, the temperature difference can reach nearly 4/K. Fig. 12(c) shows the cross-section temperature snapshot obtained from the proposed method and FEM. Fig. 12(d) shows the error snapshot comparing these two methods. The maximum temperature error is less than 0.2/K, and the errors are relatively uniformly distributed in space. The error results show that the proposed method can almost precisely predict the entire ITF.

B. Computational Time of Proposed Method

The computational time of the proposed method and the FEM are compared. The ITF predictions of the proposed method and the FEM were carried out on a desktop equipped with an Intel Core I7-1200K CPU. Table III shows the computation time required for the proposed method and the FEM. It can be seen that the main computational time of the proposed method is concentrated in the sample snapshots obtaining stage and AE-LSTM model training stage, as shown in Fig. 13. However, once the training phase is completed, the proposed method is significantly faster than the FEM. In summary, the total computation time of the proposed method is only about 20% of the FEM, which greatly improves the prediction efficiency of ITF.

TABLE IV
COMPARISON WITH EXISTING TEMPERATURE PREDICTION METHODS

Method	Computation efficiency	Temperature field	Accuracy
1-D Foster model [8]	High	No	Low
1-D Cauer model [2]	High	No	Low
3-D Foster model [33]	Medium	No	Medium
3-D Cauer model [12-13]	Medium	No	Medium
POD-LSTM model [23]	Medium	Yes	Medium
POD-ANN model [24]	Low	Yes	Medium
POD-DNN model [25]	Low	Yes	Medium
POD-BE model [26]	Medium	Yes	Medium
POD-RBF model [34]	Medium	Yes	Medium
Proposed method	High	Yes	High

C. Comparison of Prediction Methods

Table IV compares existing ITF prediction methods with the proposed method. The 1-D Foster model [8] and the 1-D Cauer model [2] can be quickly obtained through datasheets or experiments, but they are not precise due to not considering thermal coupling effects, and they can only obtain the average junction temperature. The 3-D Foster model [33] and 3-D Cauer model [12], [13] can be obtained by FEM and the modeling time increases exponentially with the model order. These two models cannot reach a high prediction accuracy due to the complex thermal coupling effects and they cannot be used for thermal analysis due to the unavailability of cross-section temperatures. The FEM can precisely predict the temperature field due to its ability to handle complex thermal coupling effects, but its prediction efficiency is low. However, the proposed method can quickly obtain the entire temperature field information and almost achieves the same prediction accuracy as the FEM due to its excellent feature extraction and learning capabilities of the hybrid model.

In addition, data-driven methods are used to accelerate finite element simulations. The current mainstream data-driven methods primarily combine POD and artificial intelligence methods. Zhou et al. [23] used POD combined with LSTM to accelerate FEM simulations. Allabou et al. [24] used POD combined with ANN to accelerate FEM simulations. Additionally, DNN [24], and BE [25], radial basis function (RBF) [34] have also been used in combination with POD to construct accelerated FEM models. The current mainstream data-driven methods have three main limitations: requires a large amount of sample data; large cumulative error in prediction; calculations take a long time. As a comparison, the proposed method decomposes sample data into stable and unstable trends, amplifying the characteristics of both trends, allowing the model to accurately extract temperature field evolution features from a small amount of sample data. This reduces the required sample data and greatly improves the computational efficiency and prediction accuracy of the proposed method.

To further verify the advantages of the proposed hybrid prediction model, it is compared with the DMD prediction model

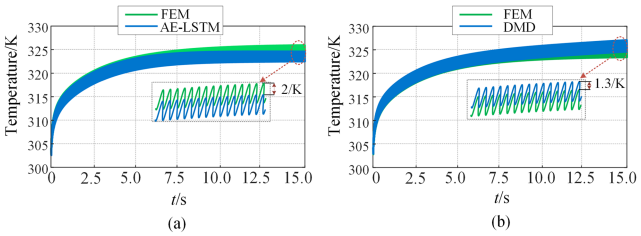

Fig. 14. Temperature prediction results at the center point of chip Q2 for different models and FEM. (a) AE-LSTM model. (b) DMD model.

TABLE V
RRMSE FOR DIFFERENT MODELS

Model	Computation time(s)	RRMSE
DMD model	3240	0.0106
AE-LSTM model	3463	0.0113
Proposed hybrid model	3464	0.0012

and the AE-LSTM prediction model. All models were tested using the parameters in Table II. The relative root-mean-square error (RRMSE) is used to evaluate the error between the FEM and the three prediction models, as

$$\text{RRMSE} = \sqrt{\frac{\| \boldsymbol{T}_{\text{FEM}} - \boldsymbol{T}_{\text{Model}} \|_2^2}{\| \boldsymbol{T}_{\text{FEM}} \|_2^2}} \tag{27}$$

where $\boldsymbol{T}_{\text{Model}}$ is the ITF snapshot obtained from the three prediction models, respectively, $\boldsymbol{T}_{\text{FEM}}$ is the ITF snapshot obtained from FEM simulation. $\|\cdot\|$ is the two-Norm of the matrix. The smaller the RRMSE, the higher the prediction accuracy of the prediction model, and vice versa.

Fig. 14 shows the temperature prediction results at the center point of chip Q2 under the FEM, AE-LSTM, DMD, respectively. It can be seen that the prediction error of the DMD model and the AE-LSTM model gradually increases with the increase of the prediction horizon. In contrast, the prediction error of the proposed model remains very small, as shown in Fig. 11(c). Table V shows the RRMSE for different models, where the RRMSE of DMD model is 0.0106, the RRMSE of AE-LSTM model is 0.0113, and the RRMSE of the proposed hybrid model is 0.0012. Thus, the prediction accuracy of the proposed hybrid model is higher than DMD model and LSTM model. Additionally, Table V shows the computation times of the three models, which are very close to each other.

D. Robustness Evaluation of Proposed Method

To verify the robustness of the proposed method to the parameters shown in Table II, three different cases are considered. In Case 1, the amplitude P_{peak} in (26) is changed. In Case 2, the amplitude f_{p} in (26) is changed. In Case 3, the sampling range S_t is changed. Table VI shows the RRMSE corresponding to Cases 1∼3.

It can be observed the RRMSE remains at the same error level in Cases 1 and 2, which indicates that the proposed method is almost unaffected by the amplitude and frequency of power loss. In Case 3, the RRMSE decreases with the increase of the sampling range S_t and maintains at a very low level. Therefore, it is recommended to increase the sampling range S_t, if more accurate prediction results are needed.

TABLE VI
RRMSE OF DIFFERENT CASES

Case	Parameters	RRMSE
Case 1	P_{peak} = 150/W	0.00115
	P_{peak} = 300/W	0.00114
	P_{peak} = 450/W	0.00118
Case 2	f_{p} = 0/Hz	0.00117
	f_{p} = 25/Hz	0.00123
	f_{p} = 100/Hz	0.00129
Case 3	S_t = 3.5s	0.00112
	S_t = 4.0s	0.00063
	S_t = 4.5s	0.00024

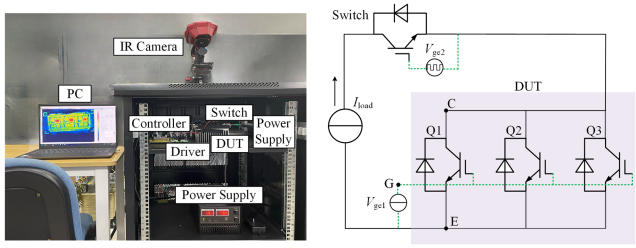

Fig. 15. Experimental platform. (a) Experimental setup. (b) Topology of experimental setup.

TABLE VII
VERIFICATION EXPERIMENT DESCRIPTION

Case	Power loss/W	Frequency/Hz	Duty ratio
I	100	0	1
II	200	5	0.5
III	300	0	1
IV	400	5	0.5

V. EXPERIMENTAL VERIFICATION

The experimental setup, as shown in Fig. 15(a), involves an IGBT module, which is colored black and serves as the device under test (DUT). This DUT is maintained in an "on" state by a voltage source. An aluminum heat sink with dimensions of 122 × 62 × 60 mm is installed under the DUT. The surface of the heat sink has dense fins to increase the heat dissipation area, and an axial fan generates a uniform airflow with a velocity of 2 m/s. A power supply is used to generate a constant current, thereby inducing power loss in the DUT. This power loss can be quantified by multiplying the current with the conduction voltage drop across the DUT. The power loss on the DUT can be modulated by applying a pulsewidth modulation (PWM) signal to the switch. An IR camera is used to capture the transient processes of the ITF. A basic power cycling test circuit topology for the IGBT was used for the experiment, as shown in Fig. 15(b).

The experiments were conducted under four operational cases, including different power losses and PWM signals with different frequencies and duty ratios, as shown in Table VII. IR camera captures and stores ITF cloud images at a sampling

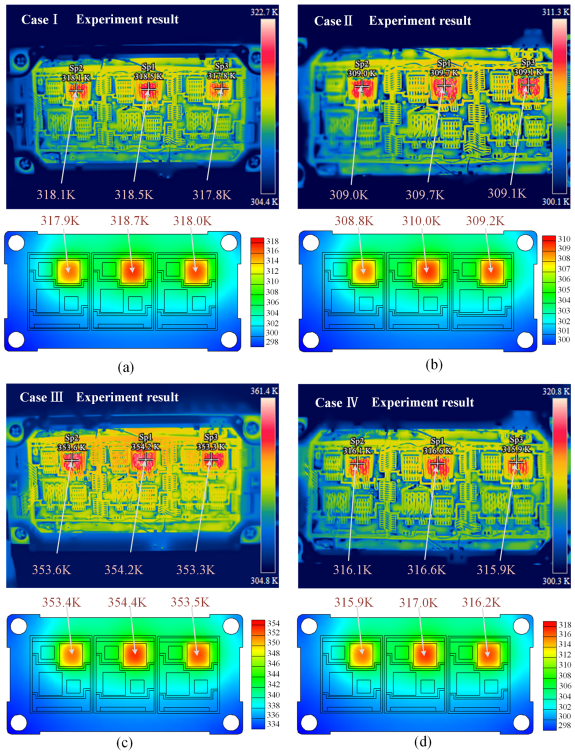

Fig. 16. Comparison results using four operational cases given in Table VII. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

TABLE VIII
MAXIMUM ERROR FOR FOUR DIFFERENT OPERATION CASES

	Case I	Case II	Case III	Case IV
Maximum error /K	0.2/K	0.3/K	0.2/K	0.4/K

frequency of 30 Hz. In the proposed method, the material properties of the heat sink and the IGBT module are the same as those used in the experiment. Air cooling is simulated by setting up a fluid domain on the surface of the heat sink and the module, with an airflow speed set to 2 m/s. The heat exchange process under forced air-cooling conditions is simulated using a fluid-structure coupling method. The model is set with boundary conditions consistent with the experiment to ensure the accuracy and credibility of the simulation results.

Fig. 16 compares ITF snapshots obtained from the experiments and the proposed method, where four cases in Table VII are considered. The ITF snapshots obtained by the proposed method are very close to the experimental results. In all four experimental results, the temperature at the center of Q1 is almost higher than that at the center of Q3, which may be due to uneven heat dissipation caused by the lack of thermal grease between the heat sink and the IGBT module. Fig. 17 shows the transient prediction results at the maximum temperature point. It can be seen that the transient error between the proposed method and the experimental results is very small, verifying the accuracy of the proposed method.

Table VIII shows the maximum absolute errors between the experimental results and the proposed method. The maximum

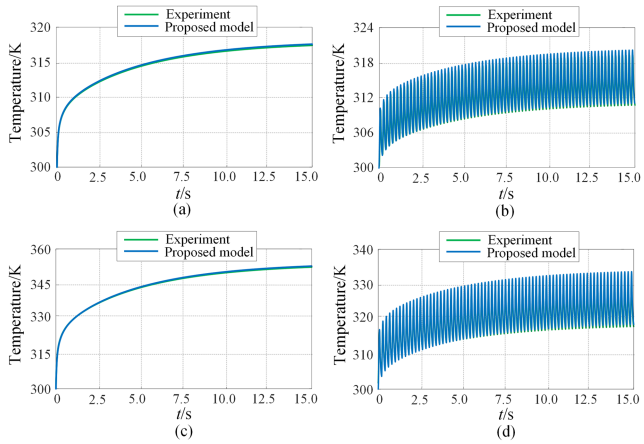

Fig. 17. Comparison of the proposed method with experimental results. (a) Case I. (b) Case II. (c) Case III. (d) Case IV.

errors under the four operational cases are 0.2/K, 0.3/K, 0.2/K, and 0.4/K, respectively. These errors could be due to the natural convective heat transfer from the surface of the open IGBT module that is exposed to the air during the experiment. This factor has not been accounted for in the proposed method's model, which might result in overestimated temperature predictions. In conclusion, the experimental data aligns closely with the predicted results.

VI. CONCLUSION

This article proposes a hybrid model prediction method based on DMD and deep learning for ITF prediction. The ITF snapshot data from FEM simulation is decomposed to extract the stable trend and the unstable trend of ITF using DMD. Then, the future stable trend is predicted by recursion and the future unstable trend is predicted by deep learning model AE-LSTM. Finally, the hybrid model successfully predicts the ITF with an acceptable error level. The proposed method almost achieves the same prediction accuracy as FEM but with only 20% of the computation time. The proposed method can obtain all the temperature field information, not just the junction temperature. It is worth mentioning that the proposed method applies not only to the ITF prediction but also to the temperature field prediction of other power electronic devices, including transformers, inductors, capacitors, silicon carbide devices, and so on.

APPENDIX
CALCULATION STEPS OF DMD METHOD

Step-1: To avoid dealing with a large-size matrix, truncated singular value decomposition [31] is applied for STDM $\boldsymbol{T}_{\mathrm{STDM}\_x}$ as

$$\boldsymbol{T}_{\mathrm{STDM}\_x} \approx \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^{*}. \tag{28}$$

Here, $\tilde{\boldsymbol{U}} \in \mathbb{C}^{m\times r}$ and $\tilde{\boldsymbol{V}} \in \mathbb{C}^{n\times r}$ are singular vector matrices, respectively. $*$ is the conjugate transpose. $\tilde{\sum} \in \mathbb{C}^{r\times r}$ is the singular value diagonal matrix, given as

$$\tilde{\boldsymbol{\Sigma}} = \mathrm{diag}[\sigma_1\sigma_2 \ldots \sigma_i \ldots \sigma_r](\sigma_1 > \sigma_2 > \ldots > \sigma_r) \tag{29}$$

where σ_i is the singular value. To improve the computational speed, the r ($1 \leq r \leq n$) largest singular values and the corresponding singular vectors are truncated. The r is recommended to be half the number of sampled snapshots n [35].

Step-2: According to (8) and (28), the dynamic matrix $\boldsymbol{S}$ can be obtained as

$$\boldsymbol{S} = \boldsymbol{T}_{\text{STDM\_}y}\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{U}}^{*}. \tag{30}$$

Step-3: The dynamic matrix $\boldsymbol{S}$ in (30) is projected onto $\tilde{\boldsymbol{U}}$ to obtain a similar low-dimensional matrix $\tilde{\boldsymbol{S}}$ given as

$$\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{U}}^{*}\boldsymbol{S}\tilde{\boldsymbol{U}} = \tilde{\boldsymbol{U}}^{*}\boldsymbol{T}_{\text{STDM\_}y}\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^{-1}. \tag{31}$$

Step-4: According to the similarity property of matrices, the dynamic matrix $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$ have the same coefficient matrix λ. Based on (31), the coefficient matrix λ of the dynamic matrix $\boldsymbol{S}$ can be obtained by a spectral decomposition of $\tilde{\boldsymbol{S}}$ given as

$$\begin{aligned}\boldsymbol{\lambda} &= \boldsymbol{W}^{-1}\tilde{\boldsymbol{S}}\boldsymbol{W} \\ &= diag[\lambda_1\lambda_2 \ldots \lambda_j \ldots \lambda_r]\end{aligned} \tag{32}$$

where $\boldsymbol{W}$ is the eigenvector matrix of $\tilde{\boldsymbol{S}}$.

Step-5: Based on (30) and (32), the modes matrix $\varphi \in \mathbb{C}^{m\times r}$ of the dynamic matrix $\boldsymbol{S}$ can be obtained as

$$\begin{aligned}\boldsymbol{\varphi} &= \boldsymbol{T}_{\text{STDM\_}y}\tilde{\boldsymbol{V}}\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{W} \\ &= [\boldsymbol{\varphi}_1\boldsymbol{\varphi}_2 \ldots \boldsymbol{\varphi}_j \ldots \boldsymbol{\varphi}_r]\end{aligned} \tag{33}$$

with

$$\boldsymbol{\varphi}_j = \begin{bmatrix}\boldsymbol{\varphi}_{j(1)} \\ \boldsymbol{\varphi}_{j(2)} \\ \ldots \\ \boldsymbol{\varphi}_{j(m)}\end{bmatrix}. \tag{34}$$

REFERENCES

[1] N. Iwamuro and T. Laska, "IGBT history, state-of-the-art, and future prospects," *IEEE Trans. Electron Devices*, vol. 64, no. 3, pp. 741–752, Mar. 2017.

[2] M. Du, Q. Guo, H. Wang, Z. Ouyang, and K. Wei, "An improved cauer model of IGBT module: Inclusive void fraction in solder layer," *IEEE Trans. Compon., Packag. Manuf. Technol.*, vol. 10, no. 8, pp. 1401–1410, Aug. 2020.

[3] J. Wang et al., "A transient 3-D thermal modeling method for IGBT modules considering uneven power losses and cooling conditions," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 4, pp. 3959–3970, Aug. 2021.

[4] X. Xiao et al., "An adaptive temperature observer for electrothermal analysis of IGBT based on temperature characteristics," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 11, no. 2, pp. 2246–2258, Apr. 2023.

[5] J. Wang et al., "A novel approach to model and analyze uneven temperature distribution among multichip high-power modules and corresponding method to respecify device SOA," *IEEE Trans. Power Electron.*, vol. 37, no. 4, pp. 4626–4639, Apr. 2022.

[6] X. Yang, S. Xu, K. Heng, and X. Wu, "Distributed thermal modeling for power devices and modules with equivalent heat flow path extraction," *IEEE Trans. Power Electron.*, vol. 11, no. 6, pp. 5863–5876, Dec. 2023.

[7] N. Baker, L. Dupont, S. Munk-Nielsen, F. Iannuzzo, and M. Liserre, "IR camera validation of IGBT junction temperature measurement via peak gate current," *IEEE Trans. Power Electron.*, vol. 32, no. 4, pp. 3099–3111, Apr. 2017.

[8] K. Ma, M. Xu, and B. Liu, "Modeling and characterization of frequency-domain thermal impedance for IGBT module through heat flow information," *IEEE Trans. Power Electron.*, vol. 36, no. 2, pp. 1330–1340, Feb. 2021.

[9] Y. Zhao, E. Deng, M. Pan, Y. Zhang, and Y. Huang, "Influence of thermal coupling on lifetime under power cycling test," *IEEE Trans. Power Electron.*, vol. 37, no. 11, pp. 13641–13651, Nov. 2022.

[10] Y. Shi, J. Liu, Y. Ai, S. Chen, and Y. Bai, "Dynamic IGBT three-dimensional thermal network model considering base solder degradation and thermal coupling between IGBT chips," *IEEE Trans. Transp. Electrific.*, vol. 9, no. 2, pp. 2994–3011, Jun. 2023.

[11] W. Guo et al., "A thermal estimation method for IGBT module adaptable to operating conditions," *IEEE Trans. Power Electron.*, vol. 36, no. 6, pp. 6147–6152, Jun. 2021.

[12] M. Ma et al., "A three-dimensional boundary-dependent compact thermal network model for IGBT modules in new energy vehicles," *IEEE Trans. Ind. Electron.*, vol. 68, no. 6, pp. 5248–5258, Jun. 2021.

[13] C. Zhan, L. Zhu, W. Wang, Q. Jiao, Y. Zhang, and S. Ji, "A large-scale identification approach for thermal parameters of multichips IGBT modules based on LLSO-SQP algorithm," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 11, no. 3, pp. 2554–2564, Jun. 2023.

[14] Z. Wang, W. Qiao, and L. Qu, "A real-time adaptive IGBT thermal model based on an effective heat propagation path concept," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 4, pp. 3936–3946, Aug. 2021.

[15] Y. Lu et al., "A 3-D temperature-dependent thermal model of IGBT modules for electric vehicle application considering various boundary conditions," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 12, no. 6, pp. 5463–5475, Dec. 2024.

[16] Y. Chen et al., "Thermal mitigation and optimization via multitier bond wire layout for IGBT modules considering multicellular electro-thermal effect," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 7299–7314, Jun. 2022.

[17] Y. Huang, Y. Luo, F. Xiao, B. Liu, and X. Tang, "Temporal and spatial differences in thermal transfer behavior of IGBTs caused by the baseplate and die-attach solder cracking," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 6, pp. 7177–7187, Dec. 2022.

[18] X. Yang, K. Heng, X. Dai, X. Wu, and G. Liu, "A temperature dependent Cauer model simulation of IGBT module with analytical thermal impedance characterization," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 3, pp. 3055–3065, Jun. 2022.

[19] C. Entzminger, W. Qiao, L. Qu, and J. L. Hudgins, "Automated extraction of low-order thermal model with controllable error bounds for SiC MOSFET power modules," *IEEE Trans. Power Electron.*, vol. 39, no. 1, pp. 538–551, Jan. 2024.

[20] Y. Zhao, Z. Wang, D. Luo, C. Chen, B. Ji, and G. Li, "Multitimescale thermal network model of power devices based on POD algorithm," *IEEE Trans. Power Electron.*, vol. 39, no. 4, pp. 3906–3924, Apr. 2024.

[21] X. Dong, A. Griffo, and J. Wang, "Multiparameter model order reduction for thermal modeling of power electronics," *IEEE Trans. Power Electron.*, vol. 35, no. 8, pp. 8550–8558, Aug. 2020.

[22] G. Fan, X. Li, and M. Canova, "A reduced-order electrochemical model of li-ion batteries for control and estimation applications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 76–91, Jan. 2018.

[23] L. Zhou et al., "High-fidelity wind turbine wake velocity prediction by surrogate model based on d-POD and LSTM," *Energy*, vol. 275, no. 15, 2023, Art. no. 127525.

[24] M. Allabou, R. Bouclier, P. Garambois, and J. Monnier, "Reduction of the shallow water system by an error aware POD-neural network method: Application to floodplain dynamics," *Comput. Methods Appl. Mechan. Eng.*, vol. 428, no. 1, 2024, Art. no. 117094.

[25] X. Zhao et al., "A hybrid method based on proper orthogonal decomposition and deep neural networks for flow and heat field reconstruction," *Expert Syst. Appl.*, vol. 247, no. 1, 2024, Art. no. 123137.

[26] M. Guo, S. McQuarrie, and K. Willcox, "Bayesian operator inference for data-driven reduced-order modeling," *Comput. Methods Appl. Mechan. Eng.*, vol. 402, no. 1, 2022, Art. no. 115336.

[27] W. Guo, M. Ma, H. Wang, H. Wang, Q. Song, and W. Chen, "A partition decoupling algorithm for compact thermal model in multichip IGBT modules," *IEEE Trans. Power Electron.*, vol. 38, no. 1, pp. 66–72, Jan. 2023.

[28] M. Xu et al., "Frequency-domain thermal coupling model of multichip power module," *IEEE Trans. Power Electron.*, vol. 38, no. 5, pp. 6522–6532, May 2023.

[29] S. Zhou et al., "Ambient data-driven online tracking of electromechanical modes using recursive subspace dynamic mode decomposition," *IEEE Trans. Power Syst.*, vol. 38, no. 6, pp. 5257–5266, Nov. 2023.

[30] A. Essien and C. Giannetti, "A deep learning model for smart manufacturing using convolutional LSTM neural network autoencoders," *IEEE Trans. Ind. Inform.*, vol. 16, no. 9, pp. 6069–6078, Sep. 2020.

[31] Z. Khatir, R. Lallemand, A. Ibrahim, and D. Ingrosso, "Thermal stress analysis comparison in IGBT power modules between DC and switching power cycling," *IEEE Trans. Power Electron.*, vol. 38, no. 9, pp. 11500–11506, Sep. 2023.
[32] Y. Zhang, H. Wang, Z. Wang, Y. Yang, and F. Blaabjerg, "Simplified thermal modeling for IGBT modules with periodic power loss profiles in modular multilevel converters," *IEEE Trans. Ind. Electron.*, vol. 66, no. 3, pp. 2323–2332, Mar. 2019.
[33] A. S. Bahman, K. Ma, and F. Blaabjerg, "A lumped thermal modelincluding thermal coupling and thermal boundary conditions for high-power IGBT modules," *IEEE Trans. Power Electron.*, vol. 33, no. 3, pp. 2518–2530, Mar. 2018.
[34] Y. Lu et al., "An efficient design method of indoor ventilation parameters for high-speed trains using improved proper orthogonal decomposition reconstruction," *J. Building Eng.*, vol. 71, no. 15, 2023, Art. no. 106600.
[35] Y. Xu et al., "A data-driven koopman approach for power system nonlinear dynamic observability analysis," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 4090–4104, Mar. 2024.

Jiahao Geng received the M.S. degree in electrical engineering from Henan Polytechnic University, Jiaozuo, China, in 2021. He is currently working toward the Ph.D. degree in electrical engineering with the School of Electrical Engineering, Southeast University, Nanjing, China.

His research interests include thermal modeling and reliability analysis of power devices, and multilevel converter control in renewable energy systems.

Fujin Deng (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the China University of Mining and Technology, Jiangsu, China, in 2005, the M.Sc. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008, and the Ph.D. degree in energy technology from the Department of Energy Technology, Aalborg University, Aalborg, Denmark, in 2012.

From 2013 to 2015 and from 2015 to 2017, he was a Postdoctoral Researcher and an Assistant Professor, respectively, with the Department of Energy Technology, Aalborg University. He was with the Southeast University in 2017 as a Professor with the School of Electrical Engineering, Southeast University, Nanjing, China. His main research interests include wind power generation, multilevel converters, high-voltage direct-current technology, dc grid, and offshore wind farm-power systems dynamics.

Qiang Yu received the B.Eng. degree from Jilin University, Changchun, China, in 2017, and Ph.D. degree from Southeast University, Nanjing, China, in 2023, both in electrical engineering.

From 2022 to 2023, he was a Visiting Ph.D. Student with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. He is currently with Southeast University as a Postdoctoral Researcher. His research interests include modular multilevel converters and their applications.

Yaqian Zhang (Member, IEEE) received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2016, and the Ph.D. degree from Southeast University, Nanjing, China, in 2023, both in electrical engineering.

From 2021 to 2022, she was a visiting Ph.D student with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. She is currently a Lecturer with Southeast University. Her research interests include medium- and high-voltage power electronic converters including modular multilevel converter and the solid-state transformer.

Zhe Chen (Fellow, IEEE) received the B.Eng. and M.Sc. degrees in electrical engineering from Northeast China Institute of Electric Power Engineering, Jilin City, China, in 1982 and 1986, respectively, and the Ph.D. degree in power and control from the University of Durham, Durham, U.K., in 1997.

He has been a Full Professor with the Department of Energy Technology, Aalborg University, Denmark, since 2002. He is the leader of Wind Power System Research team with the Department of Energy Technology, Aalborg University. He has led many research projects and has more than 1000 publications in these technical fields. His main research interests include renewable energy, power electronics, power systems, multienergy systems, and AI applications in energy systems.

Dr Chen is an Associate Editor for IEEE TRANSACTIONS ON POWER ELECTRONICS, a member of editorial boards for many international journals, a Fellow of the Institution of Engineering and Technology (IET, London, U.K.), and a Chartered Engineer in the U.K.

Frede Blaabjerg (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1995.

He was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998 with AAU Energy. In 2017, he was a Villum Investigator. He is honoris causa with University Politehnica Timisoara, Romania, in 2017 and Tallinn Technical University, Estonia, in 2018. He has authored and coauthored more than 800 journal papers in the fields of power electronics and its applications. He is the coauthor of ten monographs and editor of twenty books in power electronics and its applications e.g. the series (4 volumes) Control of Power Electronic Converters and Systems published by Academic Press/Elsevier. His current research interests include power electronics and its applications such as in wind turbines, PV systems, reliability, Power-2-X, power quality, and adjustable speed drives.

Dr. Blaabjerg was the recipient of 46 IEEE Prize Paper Awards, the IEEE PELS Distin guished Service Award in 2009, the EPE-PEMC Council Award in 2010, the IEEE William E. Newell Power Electronics Award 2014, the Villum Kann Rasmussen Research Award 2014, the Global Energy Prize in 2019, and the 2020 IEEE Edison Medal. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON POWER ELECTRONICS from 2006 to 2012. He has been Distinguished Lecturer for the IEEE Power Electronics Society from 2005 to 2007 and for the IEEE Industry Applications Society from 2010 to 2011 as well as 2017 to 2018. In 2019 and 2020, he was a President of IEEE Power Electronics Society. He was Vice-President with the Danish Academy of Technical Sciences. He is nominated in 2014–2021 by Thomson Reuters to be between the most 250 cited researchers in Engineering in the world.