Self-Adaptive RUL Prediction of Power Electronic Devices with Package Failure

Chao Guo[1] and Zhonghai Lu[2]

[1,2] *The Hong Kong University of Science and Technology (Guangzhou), China*
cguo758@connect.hkust-gz.edu.cn
zhl@hkust-gz.edu.cn

ABSTRACT

Power electronic devices vary in lifetime due to intrinsic device characteristics and extrinsic operational environments, which pose significant challenges in lifetime prediction. Traditional Deep Learning methods often directly map precursor signals to the Remaining Useful Lifetime (RUL), lacking the health state information needed to adapt dynamically to device characteristics. To address this limitation, we propose a stateful, self-adaptive RUL prediction method for package failure of power diodes. It utilizes junction temperature signals as inputs, representing thermal-mechanical fatigue influenced by external operational environments, to adjust the algorithm states, which contain the device characteristics and health state information. The proposed method combines two models, a stateful-LESIT (SLESIT) model and a Kalman Filter (KF). The SLESIT model dynamically adjusts its state using current junction temperature signals to estimate the RUL. The produced estimation is then used to rectify the predictions from an intuitive RUL propagation model in KF, providing a statistically optimal RUL estimation at each cycle. Validated through online simulation with accelerated aging data from power diodes that exhibit significant lifetime variability (68.1%), our approach reduces Mean Absolute Error (MAE) from 44.17% to 84.52% compared to popular Deep Learning methods.

1. INTRODUCTION

Power electronic devices are critical components in modern energy systems, including electric vehicles and industrial plants. Accurate prediction of their Remaining Useful Life (RUL) is essential for ensuring system reliability and reducing maintenance costs. However, the lifetime of these devices exhibits substantial variability (Kardan, Shekhar, & Bauer, 2025) due to intrinsic device characteristics (e.g., process variation) and extrinsic operational dynamics (e.g., load current variation and temperature change). This variability poses a challenge to traditional RUL prediction methods, which typically lack adaptability to device-specific degradation processes.

Existing approaches, particularly Neural Network (NN)-based methods, such as self-attention-based networks (Xiao et al., 2022), Long short-term memory (LSTM) (W. Li, Wang, Liu, Zhang, & Wang, 2020), and Recurrent Neural Network (RNN) (Cai & Lu, 2024), prefer using precursor signals or their derived features to predict RUL directly. This implies that each precursor signal sequence is labeled with an RUL value in training and inference. However, this technique suffers from a critical many-to-one mapping problem: identical precursor sequences from different devices with divergent lifetimes can map to inconsistent RUL labels due to varying device characteristics and operational environments. It forces models to converge to averages of these conditions, incurring significant prediction bias. In addition, the model parameters are static after training, thus they cannot dynamically adapt to variations in devices and operation environments.

Instead of relying only on the current input sequence to predict RUL, we propose a stateful, self-adaptive RUL prediction framework that makes predictions incorporating the health states. It preserves device characteristics and health condition information, allowing the proposed algorithm to adapt to devices with variations through state propagation. The proposed method integrates a stateful-LESIT (SLESIT) model and a Kalman Filter (KF). The SLESIT model transitions its state based on junction temperature signals, which directly relate to thermal-mechanical fatigue, and produces RUL estimations based on the state. The output of the SLESIT model is used as the observation to rectify the RUL propagation model in the KF. We employ the standard KF rather than its variants, such as the Extended Kalman Filter and Particle Filter, because it suits the linear RUL descending scenario and is capable of producing optimal estimations. In this study, we focus on the package-related wire-bond failure in power diodes.

We summarize our contributions as follows:

- We introduce a stateful framework for RUL prediction that combines SLESIT and KF, both capable of continuously updating their internal states, to adapt dynamically to variations in device characteristics and operational conditions.
- We integrate a modified lifetime model, the SLESIT model, to enhance the model interpretability. It utilizes the average junction temperature values, which are related to thermal-mechanical fatigue causing wire-bond failure, as the model states.
- To validate the effectiveness, we implement experiments against an accelerated aging dataset with variations in device characteristics and operational environments. The results show that our method outperforms the popular NN-based networks in RUL prediction, including a state-of-the-art self-attention-based network, an LSTM, and a vanilla RNN.

The remainder of this paper is structured as follows: Section 2 reviews related work; Section 3 formalizes the challenge of RUL prediction; Section 4 presents the proposed method; Section 5 evaluates the proposed method; and Section 6 concludes the work.

2. RELATED WORKS

2.1. Common failure precursors

Failure precursor signals change with the device degradation. Common failure precursors include emitter voltage V_{CE} (Haque, Choi, & Baek, 2018) and gate-emitter threshold voltage $V_{GE(th)}$ for Insulated-gate bipolar Transistors (IGBTs), on-state resistance R_{on} (R. Celaya, Saxena, Saha, & F. Goebel, 2011; Z. Li, Zheng, & Outbib, 2018), on-state voltage $V_{DS,on}$ (Vaccaro, Biadene, & Magnone, 2023), and threshold voltage V_{th} (Saha, Celaya, Vashchenko, Mahiuddin, & Goebel, 2011) for power Metal–Oxide–Semiconductor Field-Effect Transistors (MOSFETs), and forward voltage drop V_f (Lu & Otto, 2024) for power diodes. Those precursor signals are sensitive to different failure modes, thus the appropriate failure precursor should be selected for and lifetime prediction.

Forward-voltage-related signals such as V_{CE} in IGBT and V_f in power diodes are sensitive to wire-bond degradation (Hanif, Yu, DeVoto, & Khan, 2019; Smet et al., 2011; Smet, Forest, Huselstein, Rashed, & Richardeau, 2013). This is because V_{CE} increases are primarily driven by heel cracks and lift-offs in aluminum wire-bonds (Smet et al., 2013). Those failures can disconnect wire segments, reduce conductive cross-sectional area, and elevate on-state resistance and thus V_{CE}. As a result, these signals are popular in diagnosing wire-bond failure. For example, the work (Zhang et al., 2021) has linked V_{CE} drift to wire-bond shedding in IGBTs to predict the RUL based on Least Squares Support Vector Machines and Particle Filter. In SiC MOSFETs, the on-state voltage $V_{DS,on}$ has been used for training a bidirectional LSTM for RUL prediction (Vaccaro et al., 2023). For power diodes, a 5% rise in forward voltage drop V_f has been used as the end-of-life failure criterion for health condition prediction (Lu & Otto, 2024).

2.2. Data-driven methods

Deep learning models are popular in RUL prediction as they can capture nonlinear degradation patterns in precursor signals. Bidirectional LSTMs have been implemented to estimate the RUL of IGBTs encapsulated in TO-247 (Vaccaro et al., 2023). The model can leverage temporal dependencies in voltage degradation profiles, yet it demands extensive training data for accurate predictions. A self-attention (SA)-based network (Xiao et al., 2022) has been proposed for RUL prediction of IGBTs. The positional embedding and self-attention mechanism are the core structures that extract the degradation pattern for accurate prediction, with a relatively small training sample size. The LSTM has been combined with Particle Filters for IGBT failure prediction (Yang, Zhang, Li, & Miao, 2022). A data-driven approach has been proposed that employs degradation phase durations of V_{CE} as inputs to NNs and Adaptive Neuro-Fuzzy Inference System models for IGBT RUL prediction (Ahsan, Stoyanov, & Bailey, 2016). These models are limited to operate on a sequence of local failure precursors, denoted as $RUL_t = f(x_{t-n}, ..., x_{t-1})$, ignoring the variations in device characteristics and operational environments. As a result, the model outputs are biased. We propose a stateful framework that incorporates lifetime state to adapt to these variables.

2.3. Lifetime models

The Coffin-Manson law (Coffin, 1954) describes the relationship between plastic strain and fatigue, which is suitable for devices under accelerated power cycling tests. The classic lifetime model, named the LESIT model (Held, Jacob, Nicoletti, Scacco, & Poech, 1997), further combines the Coffin-Manson law with an Arrhenius term to model the wire-bond failure. It utilizes the junction temperature T_j to model the relation between lifetime and thermal fatigue. Subsequent research has enhanced the LESIT model by incorporating critical operational parameters, including conduction time t_{on} and load current density I_A for varying load conditions (Otto & Rzepka, 2019). Furthermore, a model that incorporates more device parameters, such as the diameter of bond wires, has been proposed for different types of devices (Bayerer, Herrmann, Licht, Lutz, & Feller, 2008). In this study, we focus on diagnosing a single type of device under constant loads, thus we select the traditional LESIT model. However, the critical limitation of this model lies in its static form, inherently limiting its applicability in dynamic RUL prediction. A

dynamic LESIT model (Lu, Shi, Otto, & Albrecht, 2023) has been introduced to compute the dynamic RUL by including time. We further introduce a state mechanism, evolving the model to an SLESIT model that can dynamically update its state to predict RULs during the degradation process.

3. THE RUL PREDICTION CHALLENGE UNDER VARIED LIFETIMES

3.1. The device under test and aging dataset

The diode aging dataset is sourced from paper (Lu & Otto, 2024). The power diodes are encapsulated in the power package TO-220, as shown in Figure 1. The chip size is $2.8mm \times 2.8mm$ and the wire-bonds have a diameter of $350\mu m$. The failure mode of the dataset is wire-bond lift-off failure.

Figure 1. TO-220 power package appearance (Lu & Otto, 2024).

3.2. Lifetime variation of power electronic devices

There are 6 devices in the dataset. The tested devices were repeatedly powered on/off to simulate and accelerate the degradation process. The load parameters are listed in Table 1. They have slight differences across devices: the load current I_{load} is 27-28 A, the average junction temperature swing ΔT_j is 105.1-111.5 K (6.1% variation), and the average maximum junction temperature $T_{j,max}$ is 143.4-155.0 °C (8.1% variation). However, the result lifetime N_f ranges from 10392 to 17567 cycles, with a variation of 68.1%. The highly disproportionate results imply that not only the external environment but also the intrinsic characteristics of devices can influence the lifetime. It can be further validated through a counterexample of the external environment effects. During accelerated aging testing, temperature rises when power is on (due to electrical heating) and falls when power is off. This temperature fluctuation can cause thermal-mechanical stress at the interface of wire-bonds and chips due to the mismatch in coefficients of thermal expansion. In principle, when the temperature swing and the absolute temperature change are higher, the lifetime should be shorter because of the higher thermal-mechanical stress. Nevertheless, consider the devices 2 and 5: Device 5 has both higher ΔT_j (111.5 K compared to 108.9 K) and $T_{j,max}$ (155.0 °C compared to 151.0 °C), while it has a longer lifetime (11431 compared to 10392). Taking both factors into account is essential for accurate lifetime prediction and for further reducing the bias in RUL prediction.

Table 1. Aging test load summary

Device No.	I_{load} [A]	ΔT_j [K]	$T_{j,m}$ [°C]	$T_{j,min}$ [°C]	$T_{j,max}$ [°C]	N_f [$cycle$]
1	28	106.8	90.0	36.6	143.4	11616
2	28	**108.9**	96.6	42.1	**151.0**	**10392**
3	28	106.6	91.9	38.6	145.2	13265
4	28	105.1	93.9	41.4	146.5	17467
5	28	**111.5**	99.3	43.5	**155.0**	**11431**
6	27	107.1	92.9	39.3	146.4	13507

3.3. Limitations of RUL reference curve mapping

The forward voltage drop V_f is the commonly used precursor signal as it is most sensitive to the degradation process of wire-bonds in power diodes (Hanif et al., 2019). The critical technique of data preprocessing for NN models is mapping the precursor signal with RUL labels. Figure 2 illustrates the projection of the precursor signals in the aging dataset to the reference RUL curves. The V_f curves increase slowly at first, then significantly near the end-of-life. It reflects the degradation pattern of the wire-bonds, corresponding to wire-bond crack initiation and accelerated growth. In some state-of-the-art works, they divide the whole lifetime into healthy and sub-healthy stages accordingly (Xiao et al., 2022; Lu & Otto, 2024). Each RUL reference curve is a normalized monotonic decreasing curve from 1 to 0. The x-axis intercept and the slope are dependent on the lifetime length. Each sample point on the V_f curve aligns vertically with a corresponding point on the RUL curve, ensuring that both curves are of equal length. In data preparation, the precursor signal samples or the extracted feature samples are labeled with the corresponding RUL value at the same timestamp. For example, when the V_f signal of Device 6 is around 1.677 V, it is mapped to a RUL of around 0.95.

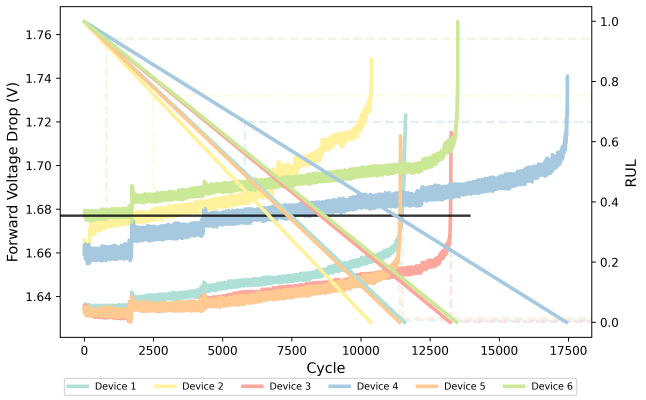


Figure 2. V_f precursor signal in the aging dataset and the corresponding RUL reference curve.

However, a significant limitation is the many-to-one mapping problem. Due to the factors introduced in the previ-

ous subsection, the V_f signals exhibit varied initial values and lengths, as depicted in Figure 2. As a result, in different devices, the same V_f input signal can map to different RUL values. For instance, the black solid horizontal line has the y-intercept of 1.677 V, and the intersections with V_f curves locate the samples in each device. The vertical dashed lines map the V_f signal with RUL values. It intersects with all 6 V_f curves at distinct parts, resulting in 6 diverse RUL labels: 0.94, 0.75, 0.66, 0.02, 0.01, and 0.01. If the NN models are trained using the multi-labeled input value of 1.677, the optimal result achieved by minimizing prediction error is the mean of the labels, which is 0.398. This limitation introduces significant prediction bias in traditional NN models that rely exclusively on a local input sequence. Some data preprocessing techniques can mitigate this issue, such as input normalization, i.e., normalizing the V_f signal or the extracted features, and the health stages separation technique mentioned above. Nonetheless, they cannot solve the core issue, the lifetime variation. As long as the lifetime varies, the many-to-one mapping RUL problem persists. Consequently, we need a model that can circumvent the RUL reference curve mapping and gather device degradation information simultaneously.

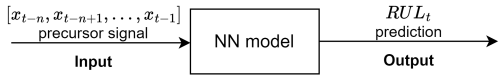
3.4. Limitations of non-adaptive models

The degradation trajectory of power electronic devices is intrinsically governed by device characteristics and operational conditions. Crucially, the fatigue accumulated per operational cycle is dynamically determined by both the applied thermal stresses and the current health state of the device. To achieve accurate RUL predictions, models must therefore continuously adapt to these evolving fatigue dynamics at each cycle. However, the parameters of non-adaptive NN models for RUL prediction remain static after training, preventing dynamic adjustments. Furthermore, those models inherently lack mechanisms to track and propagate degradation state information across cycles. These constraints can induce prediction bias, as non-adaptive models cannot capture the interplay between device-specific degradation and operational loads. Consequently, a stateful and adaptive estimation framework capable of dynamically adapting to varied device characteristics and operational conditions is essential to mitigate prediction bias.

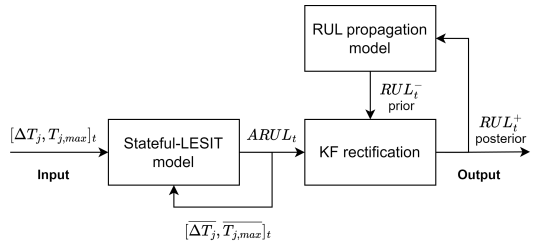
4. THE SELF-ADAPTIVE RUL PREDICTION METHOD

4.1. Overview

The proposed method adapts to the device characteristics and current health status by combining two dynamic models, the SLESIT model and KF. Conventional NN models predict failure directly from the local precursor signal sequence $f(x_{t-n}, ..., x_{t-1})$ (Figure 3a), which is indirectly related to the physics-of-failure. In contrast, the proposed model dynamically rectifies its predictions using the junction temperature signal (Figure 3b). This signal is fundamental to the wire-bond failure mechanism, indicating the health status and the fatigue endured per operational cycle. The output of SLESIT model is first corrected using the junction temperature signal and subsequently used to update the prior state estimate of the RUL propagation model via KF. The functionality and mechanism of each block will be detailed in the following.



(a) The conventional prediction framework of NN models.



(b) The self-adaptive prediction framework of the proposed method.

Figure 3. Framework comparison of NN model and proposed model.

4.2. The inherent degradation information in the junction temperature signal

The fundamental mechanism of the failure mode under study, wire-bond failure, is the thermal-mechanical stress caused by the temperature change. The difference in coefficient of thermal expansion between the Al wire-bond ($2.4 \times 10^{-5} K^{-1}$ and the Si chip ($3.0 \times 10^{-6} K^{-1}$) causes plastic deformation or even fracture in thermal cycles. Consequently, the vector $[\Delta T_j, T_{j,max}]_t$ contains the thermal fatigue information at cycle t. It results from a combination of internal device characteristics and external operational environment. When the load current is higher, the electrical thermal effect becomes more pronounced, leading to a greater temperature change. Similarly, a longer crack in the wire-bond increases its resistance, also intensifying the electrical thermal effect. Figure 4 illustrates the $T_{j,max}$ signal. Most of the devices show a rising trend under a constant load due to device degradation. This makes the signal suitable for monitoring internal device characteristics and external operational conditions to make dynamic adjustments.

In the proposed method, vector $[\Delta T_j, T_{j,max}]_t$ transmits the endured fatigue at cycle t, and the mean vector $[\overline{\Delta T_j}, \overline{T_{j,max}}]_t$ stores the health status information. In each cycle, $[\Delta T_j, T_{j,max}]_t$ adjusts the output of the SLESIT model, the Assessed RUL at cycle t ($ARUL_t$). The mean vector $[\overline{\Delta T_j}, \overline{T_{j,max}}]_t$ is calculated and reserved for the prediction of the next cycle $t + 1$.

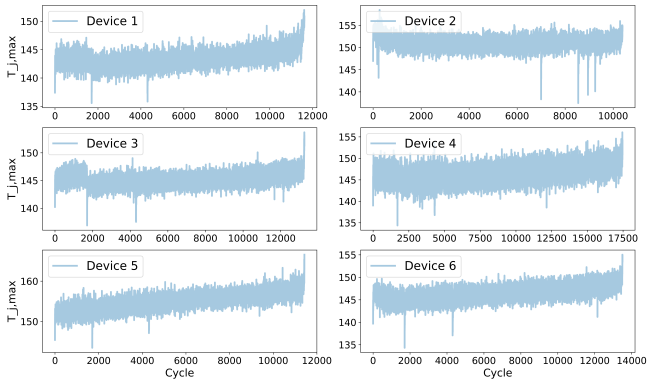


Figure 4. The $T_{j,max}$ signal of the accelerated test.

4.3. Stateful-LESIT lifetime model

4.3.1. The original and dynamic LESIT model

The LESIT model (Held et al., 1997) is an empirical model that relates junction temperature and lifetime N_f. It combines the Coffin-Manson equation and the Arrhenius term,

$$N_f = a \cdot \Delta T_j{}^{\alpha} \cdot e^{\frac{E_a}{k_B \cdot T_{j,x}}}, \tag{1}$$

where k_B is the Boltzmann constant, and $T_{j,x}$ is the absolute junction temperature. We use $T_{j,max}$ to represent the absolute junction temperature in this work. The rest parameters, a, α, and E_a, can be fitted using historical data. The dynamic LESIT model (Lu et al., 2023) includes the time t to enable dynamic calculation of RUL:

$$RUL(t) = a \cdot \Delta T_{j,t}{}^{\alpha} \cdot e^{\frac{E_a}{k_B \cdot T_{j,x,t}}} - t. \tag{2}$$

4.3.2. The Stateful-LESIT model

The original LESIT model is restricted to post-lifecycle validation of the aforementioned relationship, rather than predictive analysis. The variables ΔT_j and $T_{j,x}$ in the original LESIT model represent the average values of corresponding signals measured throughout the lifetime. When the accurate measurements of the variables are obtained, the end-of-life occurs. Consequently, the model must be modified to facilitate dynamic estimation and state propagation, ensuring compatibility with our stateful prediction framework. Based on the dynamic LESIT model, we introduce an SLESIT model that can calculate the RUL based on the previous state and the current input:

$$ARUL_t = a \cdot \left(\frac{\overline{\Delta T_{j,t-1}} \cdot (t-1) + \Delta T_{j,t}}{t}\right)^{\alpha} \cdot e^{\frac{E_a}{k_B \cdot \frac{\overline{T_{j,max,t-1}} \cdot (t-1) + T_{j,max,t}}{t}}} - t. \tag{3}$$

The input $[\Delta T_j, T_{j,max}]_t$ carries the recent thermal fatigue information of cycle t to rectify the state $[\overline{\Delta T_j}, \overline{T_{j,max}}]_{t-1}$. The rectified state representing the new degradation level at current state t is

$$[\overline{\Delta T_j}, \overline{T_{j,max}}]_t = \frac{1}{t}((t-1) \cdot [\overline{\Delta T_j}, \overline{T_{j,max}}]_{t-1} + [\Delta T_j, T_{j,max}]_t). \tag{4}$$

The $ARUL_t$ is calculated based on N_f estimation of the current state minus the past cycles t.

4.4. The Kalman Filter for self-adaptive prediction

4.4.1. The common Kalman Filter equations

KF is a recursive estimation algorithm renowned for its ability to optimally combine uncertain predictions with noisy observations to produce statistically optimal state estimates in dynamic systems. This characteristic makes it suitable for our application scenario. The basic KF equations include the state transition equation and the observation equation:

$$X_t = f(X_{t-1}) + Q_t, \tag{5}$$

$$Y_t = h(X_t) + R_t, \tag{6}$$

where X is the state variable, Y is the observation variable, $f(\cdot)$ and $h(\cdot)$ are state transition function and observation function, and Q and R are zero-mean process and observation noise variables respectively. Each variable is a random variable subject to a Gaussian distribution, which can be uniquely represented by its mean and variance. The recursive algorithm employs a predict-update loop to make estimations and rectify them iteratively. We employ this iteration in rectifying the estimation of an RUL propagation model with $ARUL_t$.

4.4.2. The Kalman Filter combined with SLESIT model

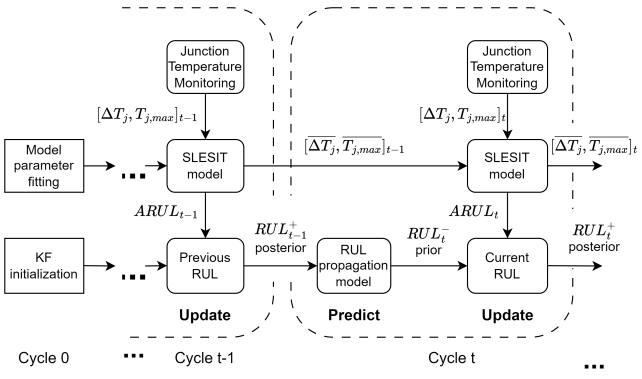


Figure 5. The KF predict-update iteration framework in self-adaptive RUL prediction.

The predict-update iteration in self-adaptive prediction is shown in Figure 5. It continuously propagates the state estimate (prior) in the predict stage, then rectifies this estimation (updating to the posterior) by incorporating the observation via the Bayesian equation. The Bayesian equation for rectifying the prior can be expressed as:

$$posterior = \frac{likelihood \times prior}{normalization\ factor}. \tag{7}$$

We denote the prior and posterior with $^-$ and $^+$, respectively. In our implementation, the state and observation variables are RUL and $ARUL$. We utilize an intuitive RUL propagation model as the state transition equation:

$$RUL_t = RUL_{t-1} - \delta(1). \tag{8}$$

The RUL distribution subtracts a Gaussian approximation of delta function with a mean of 1. It indicates that the RUL variable continuously decreases by 1 in each predict stage. The observation function is $h(x) = x$, as $ARUL$ directly rectifies RUL.

Since the Gaussian distribution calculation can be separated into mean and variance calculations, the RUL in the predict-update loop is propagated as:

- Predict:

$$RUL_t^- = RUL_{t-1}^+ - 1, \tag{9}$$
$$Var_t^- = Var_{t-1}^+ + Q_{v,t}, \tag{10}$$

  where Var_t^- is the RUL prior variance at cycle t, and $Q_{v,t}$ is the variance of the process noise.

- Update:

$$K_t = \frac{Var_t^-}{Var_t^- + R_{v,t}}, \tag{11}$$
$$RUL_t^+ = RUL_t^- + K_t(ARUL_t - RUL_t^-), \tag{12}$$
$$Var_t^+ = (1 - K_t)Var_t^-, \tag{13}$$

  where K_t is the Kalman gain at cycle t, and $R_{v,t}$ is the variance of the observation noise.

In the experiments, $Q_{v,t}$ is set to 10^{-4} and $R_{v,t}$ is set to 10^8.

5. EVALUATION

5.1. Experiment

We evaluate the real-time prediction performance with offline accelerated aging data. The experiment flow is illustrated in Figure 6. The model parameters are fitted first. At each timestamp, only the current signal value is input to the model, while all others are masked. Then it predicts the current RUL based on the state and the current input. The prediction performance is evaluated using Mean Absolute Error (MAE).

5.2. Result of proposed prediction method

The model parameters $[a, \alpha, \frac{E_a}{k_b}]$ after fitting is $[2.82 \times 10^{40}, -14.10, -7463.76]$. The RUL prediction result of the proposed self-adaptive method is shown in Figure 8. To demonstrate the effectiveness of the KF, the SLESIT model prediction result is also illustrated for comparison. The average MAE of 6 devices is presented in Table 2. Comparing the two methods:

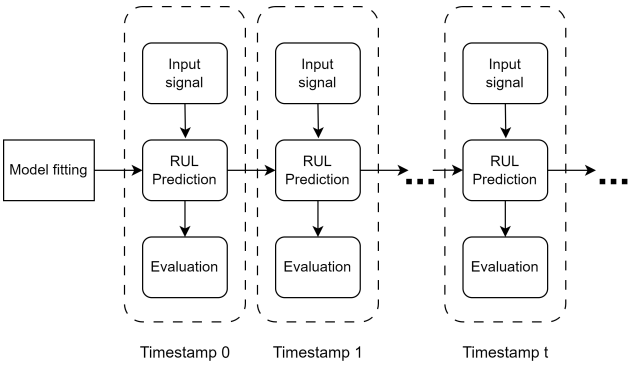


Figure 6. Evaluation experiment process.

1. The proposed method outperforms the SLESIT model, achieving an MAE of 500.82 compared to 621.75. It shows a 19.45% improvement in MAE, validating the contribution of KF.
2. The predictions of the two methods are deviated from the truth at the beginning, then converge to the true RUL reference curve. The reason is that the input samples are insufficient to get accurate estimations. When the state is initialized far from the true state, the state of the model gradually converges to the RUL reference curve. It demonstrates the adaptability of the proposed model and the SLESIT model.
3. Empirically, this convergence occurs within the first 1% of the data length. The initial state for both models is set using the first cycle measurements of junction temperature, as determined by Equation 3. Excluding the first 1% samples, the two methods against the remaining 99% prediction samples achieve MAE of 488.45 and 617.19, respectively.
4. The incorporation of KF enables the calculation of 95% confidence interval (95% CI), as illustrated in Figure 8. As the state propagates, the CI narrows and converges, exhibiting the adaptability of the proposed model.

5.3. Comparison with benchmark methods

To present the limitations of non-adaptive models under varied device characteristics and operation conditions, we implemented several NN-based methods, including a state-of-the-art method using a self-attention-based (SA) network (Xiao et al., 2022), an LSTM-based method, and an RNN-based method. The algorithm flow chart is shown in Figure 7. The red dense arrows represent the SA algorithm process flow, while the blue dashed arrows represent the LSTM/RNN algorithm flow. They utilize the most commonly used forward voltage drop precursor signal V_f for prediction. In NNs that

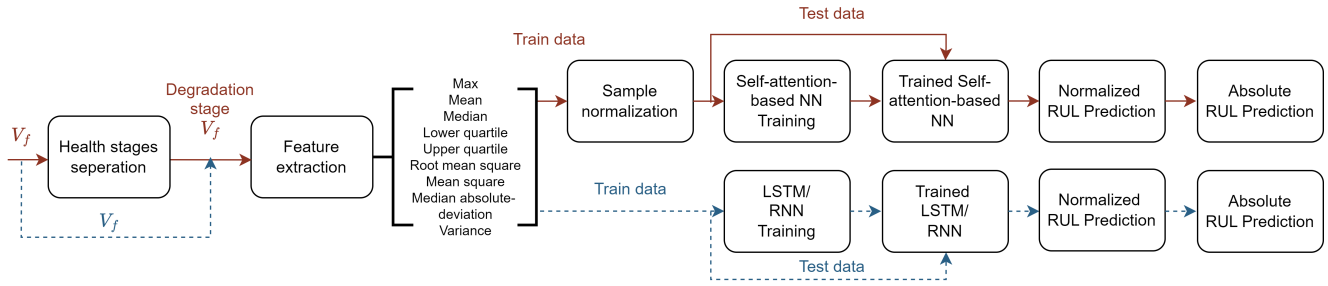


Figure 7. The flow chart of benchmark NN methods.

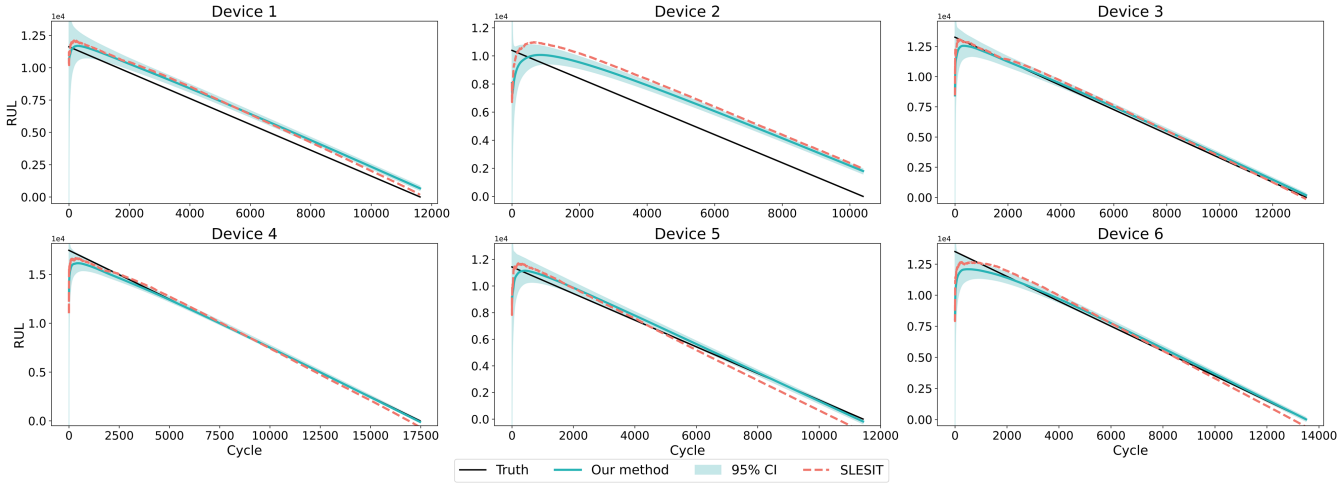


Figure 8. The prediction results of the self-adaptive prediction method and the SLESIT model.

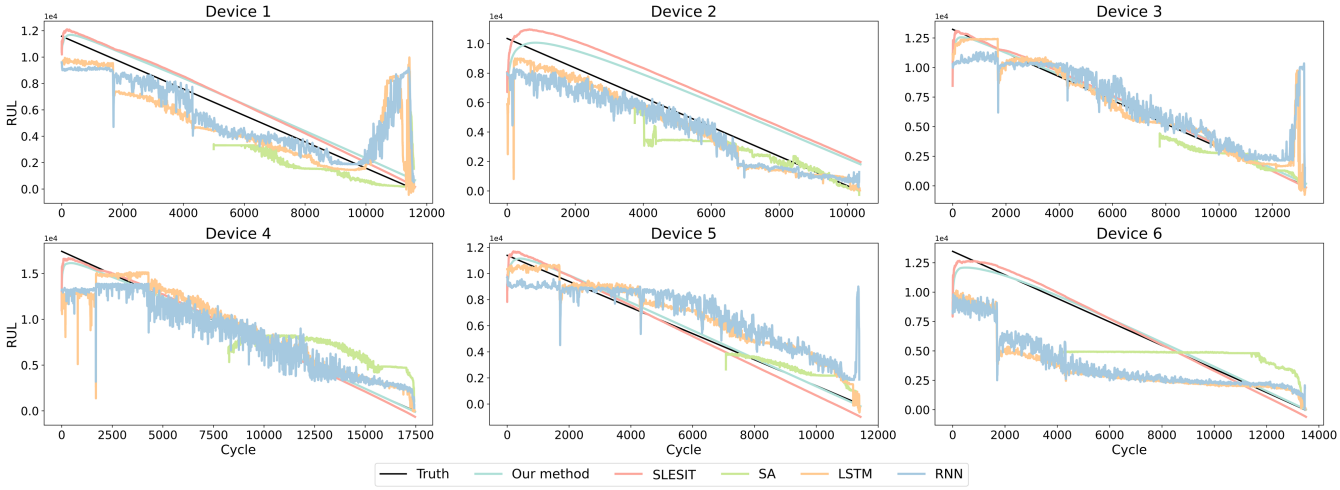


Figure 9. Prediction results of benchmark methods and the proposed method.

Table 2. Model Comparison by MAE (cycle)

Metric	**Our method**		**SLESIT**		**SA** (Xiao et al., 2022)		**LSTM**		**RNN**	
	All	99%	All	99%	Train	Test	Train	Test	Train	Test
RUL MAE	500.82	488.45	621.75	617.19	874.88	1849.28	1431.62	2196.59	1299.76	3154.59

Table 3. Network parameters of benchmark models

Parameter	**SA**	**LSTM**	**RNN**
Input features	9 time-domain features of V_f		
Learning rate	0.001		
Input (window × feature)	20 × 9 + Positional encoding	20 × 9	20 × 9
Network layers	*Multi-head self-attention:* dimension = 24	*LSTM:* 64 units	*RNN:* 64 units
	Fully-connected: [128, 64, 32, 1]	*Fully-connected:* [16, 4, 1]	*Fully-connected:* [16, 4, 1]

Table 4. Train-test device split for 6-fold experiment

Case	Train device No.	Test device No.
1	1,2,3,4,5	6
2	2,3,4,5,6	1
3	1,3,4,5,6	2
4	1,2,4,5,6	3
5	1,2,3,5,6	4
6	1,2,3,4,6	5

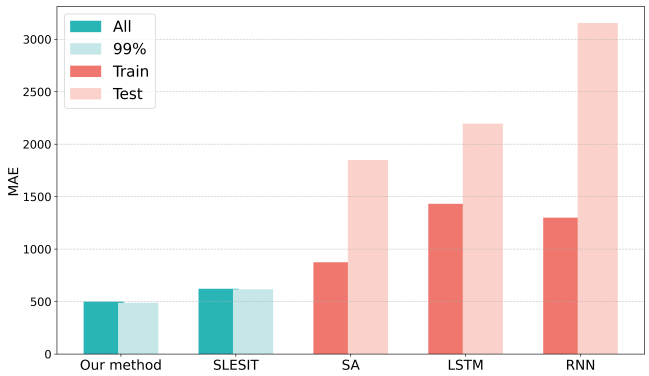


Figure 10. Model Comparison by MAE.

depend on time series signals with strong degradation patterns for prediction, the V_f signal (Figure 2) is more effective than the junction temperature signal (Figure 4) as the V_f signal has a more significant degradation trend. They both apply the fundamental RUL reference curve mapping to label the input sequence. The SA network tries to address the aforementioned lifetime variation problem and the RUL many-to-one mapping problem using a health stages separation and sample normalization technique. To mitigate the many-to-one mapping problem, the degradation state from the second half of the lifecycle is isolated for prediction. Then the length of the degradation state data is normalized to 10000 cycles by downsampling or interpolation, which avoids the lifetime variation problem. Nine time-domain features are extracted and are input to the networks for training and testing. The predicted RULs are converted to real RULs for evaluation and comparison.

We implemented a 6-fold cross-validation technique to comprehensively evaluate the NN benchmark models, using 5 devices for training and 1 device for testing in each fold. The train-test device grouping is listed in Table 4. The result of Case 1 (Device 6 for testing) is illustrated in Figure 9, and the average training and testing loss of 6 cases are concluded in Table 2 and Figure 10.

The results show that:

1. Our method achieves MAE reductions of 386.43, 943.16, and 811.30 (44.17%, 65.88%, 62.42%) and 1360.82, 1708.13, and 2666.14 (73.59%, 77.76%, 84.52%) against benchmark method (SA, LSTM, and RNN) training and testing performance, demonstrating substantial improvements across both sets.
2. The health stages separation and sample normalization technique used in the SA model improves the prediction performance. However, the health stages separation technique limits the RUL prediction range to the sub-healthy stage. In addition, normalizing the sample size to a constant needs information about the true lifetime, which is not applicable in real online prediction.
3. By incorporating the health state information, the proposed method and SLESIT generate smoother and more realistic RUL prediction curves than benchmark modes.

6. CONCLUSION

This study proposes a self-adaptive framework for predicting the RUL of power diodes, mitigating prediction bias induced by intrinsic device characteristics and extrinsic operating conditions. Building on the established LESIT and dynamic LESIT model, we introduce a stateful variant, the SLESIT model. It incorporates a state mechanism that enables it to dynamically adjust its state and estimate RUL based on junction temperature signals, which are directly linked to package failure and wire-bond fatigue. This RUL estimate is used to rectify the prior state in an RUL propagation model via a KF, yielding statistically optimal prediction results. By integrat-

ing two stateful models, the proposed method adapts dynamically to diverse device parameters and operational environments. Validation on diode aging data shows that the framework reduces MAE by 44.17% to 84.52% compared to popular NN models, including the self-attention-based model, LSTM, and RNN.

In future work, we will extend our approach to other failure modes and devices by integrating suitable physics-of-failure models, while also benchmarking against more recent data-driven and hybrid RUL methods, such as Transformer-based architectures and physics-informed neural networks.

REFERENCES

Ahsan, M., Stoyanov, S., & Bailey, C. (2016, May). Data driven prognostics for predicting remaining useful life of IGBT. In *2016 39th International Spring Seminar on Electronics Technology (ISSE)* (pp. 273–278). Pilsen, Czech Republic.

Bayerer, R., Herrmann, T., Licht, T., Lutz, J., & Feller, M. (2008, March). Model for Power Cycling lifetime of IGBT Modules - various factors influencing lifetime. In *5th International Conference on Integrated Power Electronics Systems* (pp. 1–6).

Cai, C., & Lu, Z. (2024, May). RUL Estimation for Power Electronic Devices Using RNNs. In *2024 Prognostics and System Health Management Conference (PHM)* (pp. 319–327). Stockholm, Sweden.

Coffin, L. F. (1954, August). A Study of the Effects of Cyclic Thermal Stresses on a Ductile Metal. *Journal of Fluids Engineering*, *76*(6), 931–949.

Hanif, A., Yu, Y., DeVoto, D., & Khan, F. (2019, May). A Comprehensive Review Toward the State-of-the-Art in Failure and Lifetime Predictions of Power Electronic Devices. *IEEE Transactions on Power Electronics*, *34*(5), 4729–4746.

Haque, M. S., Choi, S., & Baek, J. (2018, March). Auxiliary Particle Filtering-Based Estimation of Remaining Useful Life of IGBT. *IEEE Transactions on Industrial Electronics*, *65*(3), 2693–2703.

Held, M., Jacob, P., Nicoletti, G., Scacco, P., & Poech, M.-H. (1997). Fast power cycling test of IGBT modules in traction application. In *Proceedings of Second International Conference on Power Electronics and Drive Systems* (Vol. 1, pp. 425–430). Singapore.

Kardan, F., Shekhar, A., & Bauer, P. (2025). Impact of Parameter Uncertainties on Power Electronic Device Lifetime Predictions. *IEEE Access*, *13*, 100479–100491.

Li, W., Wang, B., Liu, J., Zhang, G., & Wang, J. (2020, November). IGBT aging monitoring and remaining lifetime prediction based on long short-term memory (LSTM) networks. *Microelectronics Reliability*, *114*, 113902.

Li, Z., Zheng, Z., & Outbib, R. (2018, September). A prognostic methodology for power MOSFETs under thermal stress using echo state network and particle filter. *Microelectronics Reliability*, *88-90*, 350–354.

Lu, Z., & Otto, A. (2024, June). Health Condition Estimation for Discrete Power Electronic Devices under Package Failure. In *2024 IEEE International Conference on Prognostics and Health Management (ICPHM)* (pp. 336–347). Spokane, WA, USA.

Lu, Z., Shi, R., Otto, A., & Albrecht, J. (2023, May). RUL Estimation for Power Electronic Devices Based on LESIT Equation. In *2023 Prognostics and Health Management Conference (PHM)* (pp. 47–54). Paris, France: IEEE.

Otto, A., & Rzepka, S. (2019). Lifetime modelling of discrete power electronic devices for automotive applications. In *AmE 2019 - Automotive meets Electronics; 10th GMM-Symposium* (pp. 1–6).

R. Celaya, J., Saxena, A., Saha, S., & F. Goebel, K. (2011, September). Prognostics of Power MOSFETs under Thermal Stress Accelerated Aging using Data-Driven and Model-Based Methodologies. *Annual Conference of the PHM Society*, *3*(1).

Saha, S., Celaya, J. R., Vashchenko, V., Mahiuddin, S., & Goebel, K. F. (2011, May). Accelerated aging with electrical overstress and prognostics for power MOSFETs. In *IEEE 2011 EnergyTech* (pp. 1–6). Cleveland, OH, USA.

Smet, V., Forest, F., Huselstein, J.-J., Rashed, A., & Richardeau, F. (2013, July). Evaluation of Vce Monitoring as a Real-Time Method to Estimate Aging of Bond Wire-IGBT Modules Stressed by Power Cycling. *IEEE Transactions on Industrial Electronics*, *60*(7), 2760–2770.

Smet, V., Forest, F., Huselstein, J.-J., Richardeau, F., Khatir, Z., Lefebvre, S., & Berkani, M. (2011, October). Ageing and Failure Modes of IGBT Modules in High-Temperature Power Cycling. *IEEE Transactions on Industrial Electronics*, *58*(10), 4931–4941.

Vaccaro, A., Biadene, D., & Magnone, P. (2023). Remaining Useful Lifetime Prediction of Discrete Power Devices by Means of Artificial Neural Networks. *IEEE Open Journal of Power Electronics*, *4*, 978–986.

Xiao, D., Qin, C., Ge, J., Xia, P., Huang, Y., & Liu, C. (2022, March). Self-attention-based adaptive remaining useful life prediction for IGBT with Monte Carlo dropout. *Knowledge-Based Systems*, *239*, 107902.

Yang, J., Zhang, H., Li, L., & Miao, Q. (2022, December). IGBT modules fault prediction based on particle filter with an improved nonlinear characteristics representation of state-space model. *Microelectronics Reliability*, *139*, 114795.

Zhang, J., Hu, J., You, H., Jia, R., Wang, X., & Zhang, X. (2021, June). A remaining useful life prediction

method of IGBT based on online status data. *Microelectronics Reliability*, *121*, 114124.

BIOGRAPHIES

Chao Guo received the B.Sc. degree in electronic science and technology from Beijing Institute of Technology, Beijing, China, in 2021, and the M.Sc. degree in electrical engineering from KTH Royal Institute of Technology, Stockholm, Sweden, in 2024. He is currently working toward the Ph.D. degree in microelectronics at Hong Kong University of Science and Technology (Guangzhou), China.

Zhonghai Lu received the BS degree in radio and electronics from Beijing Normal University, Beijing, China, in 1989, and the MS degree in systems-on-chip design and the PhD degree in electronic and computer systems design from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2002 and 2007, respectively. He was an engineer in the area of electronic and embedded systems from 1989 to 2000. He was Professor with KTH Royal Institute of Technology, and is now Professor with The Hong Kong University of Science and Technology (Guangzhou), China.