

ABSTRACT

Title of dissertation: PROGNOSTIC MODELING FOR
RELIABILITY PREDICTIONS OF
POWER ELECTRONIC DEVICES

Yizhou Lu, Doctor of Philosophy, 2019

Dissertation directed by: Professor Aris Christou
Department of Mechanical Engineering

The applications of semiconductor power electronic devices, including power and RF devices, in industry have stringent requirements on their reliability. Power devices are subject to various types of failure mechanisms under various stressors. Prognostics and health management (PHM) allows detecting signs of failures, providing warnings of failures in advance, and performing condition-based maintenance. There is a pressing need to develop a robust prognostic model to detect anomalous behavior and predict the lifetime of devices that can be applicable to different types of power transistors. In the present dissertation, a comprehensive prognostic model for remaining useful life (RUL) prediction of semiconductor power electronic devices is developed. The model consists of an anomaly detection module and a RUL prediction module including a non-linear system process model describing the evolution of parametric degradation, and a measurement model. The anomaly detection module uses principal component analysis (PCA) for dimensionality reduction and feature extraction, as well as k-means clustering to establish baseline clusters in the feature

space. The novel singular-value-weighted distance (SVWD) is developed as the distance measure in the feature space, based on which Fisher criterion (FC) is used for anomaly probability calculation. The system process model incorporates variables concerning loading conditions and physics-of-failure of devices, and uses particle filter (PF) approach for process model training and RUL prediction. For PF methodology, a novel resampling technique, called MHA-replacement resampling, is developed to solve the particle degeneracy in classic PF techniques and sample impoverishment in traditional resampling techniques. The developed prognostic model is first implemented on IGBT modules for validation. It was reported that the module package of power transistors was susceptible to various types of fatigue-related failure modes due to coefficient of thermal expansion (CTE) mismatches under temperature/power cycles introducing thermomechanical stresses. The physics-of-failure “driving variable” is derived from Paris equation. The model is validated on several time-series IGBT module degradation data under power cycles from literature sources, based on SIR particle filter for RUL prediction with good accuracy. Then the model is implemented on GaN HEMTs, a representative of wide-bandgap semiconductor power devices. GaN HEMTs are susceptible to degradation mechanisms such as ohmic contact inter-diffusion that leads to voiding in the field plate at high temperature under RF accelerated life tests (ALTs). The time-series data of the physics-of-failure “driving variable” is obtained from diffusion computation in Mathematica with the temperature profile coming from COMSOL thermal simulation. The RUL prediction results based on SIR filter are also satisfactory for GaN HEMTs. For each type of device, the new resampling technique is validated through performance benchmarking

against state-of-the-art resampling techniques. Another reliability threat for GaN HEMTs, especially in aerospace and nuclear applications, is the degradation due to radiation effect on the device performance. Gamma radiation has been found to lead to generation of defects in AlGaN/GaN layers, which form complexes acting as carrier traps, thus reducing carrier density and current. EPC GaN HEMTs are irradiated under a wide range of Gamma ray doses and critical DC characteristics are recorded before and after radiation to quantify their shifts during the irradiation. Future work needed to allow implementation of the developed prognostic model for RUL estimation is proposed.

PROGNOSTIC MODELING FOR RELIABILITY PREDICTIONS
OF POWER ELECTRONIC DEVICES

by

Yizhou Lu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:

Professor Aris Christou: *Advisor, Chair*

Professor James D. Baeder: *Dean's Representative*

Professor F. Patrick McCluskey

Professor Peter A. Sandborn

Professor Mohammad Modarres

© Copyright by
Yizhou Lu
2019

Dedication

To my parents who have inspired and supported me throughout the duration of graduate school

Acknowledgments

I would like to give my most sincere and special gratitude to my advisor, Prof. Aris Christou, for his continuous support and mentoring throughout my graduate studies. He offered me the chance to address a challenging but interesting topic. He has been supportive on any problems that I ran into on my coursework and research work. He provided the vision, inspiration, scientific advice and encouragement for me to conquer all the difficulties and complete my PhD researches.

I would like to express my gratitude to Dr. Michael Azarian and Dr. Diganta Das for their suggestions and assistance on my research.

I owe the special gratitude and appreciation to Dr. David Shahin for numerous help with my research work.

I am also very grateful to and appreciate the current and former graduate students and staff of CALCE and UMD FabLab for all the generous guidance and kind assistances provided for my lab work. I also acknowledge and extend my gratitude to the staff in UMD Radiation Facilities for the assistance in the radiation part of my researches.

I would like to thank Prof. Baeder, Prof. Sandborn, Prof. McCluskey and Prof. Modarres for serving on my dissertation committee and providing me valuable advice.

I also would like to appreciate all the MSE and ME staff in UMD for their kind and generous support when necessary.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	vii
List of Figures	viii
1 Introduction	1
1.1 Problem Statement	1
1.2 Overview of Prognostics and Health Management	4
1.3 Fault Detection	6
1.3.1 Introduction	6
1.3.2 Model-based and Data-driven Fault Detection	7
1.3.3 Supervised, Unsupervised and Semi-supervised Fault Detection	8
1.3.4 Anomaly Detection	10
1.4 Dynamic State Estimation and Particle Filters	12
1.4.1 Bayesian Filters for Dynamic State Estimation	12
1.4.2 Particle Filters	13
1.4.3 Literature Review of Particle-Filter-Based Prognostics	14
1.5 Power Electronics and IGBTs	22
1.5.1 Power Converters and Power Electronics	22
1.5.2 Insulated Gate Bipolar Transistor (IGBT) Devices and Modules	24
1.5.3 IGBT Module Package	26
1.5.4 Package-related Failure Mechanisms	27
1.5.5 Accelerated Life Testing and Power Cycling Test	30
1.6 HF Gallium Nitride (GaN) High Electron Mobility Transistors (HEMTs)	31
1.6.1 GaN Devices	31
1.6.2 HEMTs	33
1.6.3 High Frequency Operation of GaN HEMTs	34
1.6.4 Failure Mechanisms of HVHF GaN HEMTs	35
1.6.5 HF-Accelerated Life Testing (ALT)	38
1.7 Dissertation Scope and Outline	39

2	Particle Filter Approach	42
2.1	Non-linear Bayesian Filtering	42
2.2	Particle Filters	43
2.3	Sampling Importance Resampling (SIR) Filter	45
2.4	Summary	47
3	Prognostic Model Development	48
3.1	Failure Precursors and Failure Thresholds	48
3.2	Model Framework	50
3.3	Anomaly Detection Technique	51
3.3.1	Feature Space	51
3.3.2	Nominal Region	52
3.3.3	Feature Construction and Extraction	54
3.3.4	Singular-Value-Weighted Distance	56
3.3.5	K -means Clustering	58
3.3.6	Anomaly Probability Calculation	59
3.3.7	Validation of Anomaly Detection Technique	62
3.4	Bayesian Filter Model	66
3.4.1	Physics-of-Failure-Based Model Variables for IGBT Modules	66
3.4.2	Physics-of-Failure-Based Model Variables for GaN HEMTs	70
3.4.3	Model Specification	81
3.5	Summary	86
4	Novel Resampling Technique Development	89
4.1	Introduction of Resampling	89
4.2	Traditional Resampling Techniques	91
4.2.1	Multinomial Resampling	91
4.2.2	Systematic Resampling	92
4.2.3	Stratified Resampling	92
4.2.4	Residual Resampling	93
4.3	Other Resampling Techniques	97
4.4	Novel Resampling Technique: IMHA-Replacement Resampling	101
4.5	Summary	105
5	Validation of Prognostic Model for IGBT Modules	106
5.1	Degradation Data Sources	106
5.2	Results of Anomaly Detection	107
5.3	Results of RUL Estimation Based on SIR Filter	112
5.4	Resampling Techniques Implementation and Validation	120
5.5	Summary	127
6	Validation of Prognostic Model for GaN HEMTs	128
6.1	Degradation Data Source	128
6.2	Results of Anomaly Detection	129
6.3	Results of remaining useful life (RUL) Estimation Based on SIR Filter	135

6.4	Resampling Techniques Implementation and Validation	141
6.5	Summary	148
7	Conclusions and Future Work	149
7.1	Conclusions and Contributions of The Dissertation	149
7.2	Discussions and Future Work	153
A	Gamma Radiation on Enhancement-mode GaN HEMTs	156
A.1	Introduction	156
	A.1.1 Gamma Radiation	156
	A.1.2 Enhancement-mode Devices	157
	A.1.3 Radiation-induced Degradation of GaN HEMTs	158
A.2	Devices Under Test (DUTs)	159
A.3	Condition and Procedure of Gamma Radiation	160
A.4	Electrical Characterization	161
A.5	Results of Device Characteristics Pre- and Post-Irradiation	161
A.6	Hypothesis of Degradation Mechanisms and Potential for Prognostic Model Implementation	174
A.7	Summary	176
	Bibliography	177

List of Tables

1.1	Physical properties of important semiconductors for power devices [1]	32
3.1	Results of adjusted Rand index (ARI) Comparison for the proposed anomaly detection technique	66
3.2	Summary of self-diffusion coefficients of interdiffusion components . .	75
5.1	Optimal number of clusters and maximum silhouette values of IGBT module degradation data series	107
5.2	Optimal number of clusters and maximum silhouette values of IGBT module degradation data series	111
5.3	Points for onset of test data and anomaly signaling point (ASP) of IGBT module degradation data series	112
5.4	Summary of RUL prediction results for IGBT modules	119
6.1	Optimal number of clusters and maximum silhouette values of Wolf-speed GaN HEMT degradation data series	134
6.2	Optimal number of clusters and maximum silhouette values of Wolf-speed GaN HEMT degradation data	135
6.3	Summary of Future State prediction results for GaN HEMTs	141
A.1	Summary of mean variations of DC parameters for all groups of devices	172

List of Figures

1.1	A complete prognostics and health management (PHM) cycle [2] . . .	5
1.2	Generator topology and control of a doubly fed induction generator (DFIG) system [3]	23
1.3	Back-to-back converter of a wind turbine with paralleled half-bridges in each phase module to provide the required current capacity [4] . .	23
1.4	Circuit diagram of Infineon FS20R06W1E3 IGBT module [5]	25
1.5	Module Package Structure for IGBTs [5]	27
1.6	Crack growing closely above the bond wire/chip interface, leading to bond wire lift-off [6]	29
1.7	The Schematic of an AlGa _N /Ga _N heterojunction showing the formation of 2DEG layer [7]	34
3.1	Example of I-V output characteristics curves of an IGBT in Infineon FS20R06W1E3, given a gate voltage V_{GE} of 9 V and 11 V, respectively	53
3.2	Time-series data of $\Delta K(k)$ for degradation data series in [8]	70
3.3	Source contact stacked layers and source-connected field plate of Wolfspeed Ga _N HEMTs	71
3.4	Temperature profile simulation of Wolfspeed Ga _N HEMT under RF-ALT	73
3.5	Au concentration profile in the Ti/Al/Ti layers of the Ga _N HEMT source contact vs. time	79
3.6	The total number of Au atoms in the Ti/Al/Ti layers of the Ga _N HEMT source contact vs. the elapsed test time	80
4.1	Pseudo-code of multinomial, systematic and stratified resampling [9] .	93
4.2	Pseudo-code of residual resampling [9]	94
4.3	Effective sample size N_{eff} vs. cycles for "the classic four" techniques in long-term dynamic state estimation showing increasingly frequent resampling and weight reset	95
4.4	Pseudo-code of IMHA resampling [10]	99
4.5	Flow chart of the IMHA-replacement resampling technique	104
5.1	Plots of healthy data instances in the PC-based feature space including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only	108

5.2	Variation of silhouette value with the number of clusters for IGBT module degradation data series in [8]	109
5.3	Plots of baseline clusters for healthy data instances including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only	110
5.4	(a) RUL prediction results based on SIR filter; (b) Predicted cycles-to-failure (CTF) distribution of particles for the IGBT module degradation data series in [8]	114
5.5	(a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series in [11]	116
5.6	(a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series 1E in [12]	117
5.7	(a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series 2E in [12]	118
5.8	Time-series data of N_{eff} for the IGBT module degradation data series in [8]	121
5.9	Time-series data of N_{eff} for the IGBT module degradation data series 1E in [12]	122
5.10	Time-series data of N_{eff} for the IGBT module degradation data series 2E in [12]	122
5.11	Time-series data of N_{iden} for the IGBT module degradation data series in [8]	124
5.12	Time-series data of N_{iden} for the IGBT module degradation data series 1E in [12]	124
5.13	Time-series data of N_{iden} for the IGBT module degradation data series 2E in [12]	125
5.14	(a) RUL distribution based on the multinomial resampling; (b) RUL distribution based on the IMHA-replacement resampling; for the IGBT module degradation data series 1E in [12]	126
6.1	Plots of healthy data instances in the PC-based feature space including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only	130
6.2	Variation of silhouette value with the number of clusters for Device 4 degradation data series	132
6.3	Plots of baseline clusters for healthy data instances including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only	133
6.4	(a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at t=800hrs for the degradation data series of Device 1	136
6.5	(a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at t=800hrs for the degradation data series of Device 2	138

6.6	(a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at t=800hrs for the degradation data series of Device 3	139
6.7	(a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at t=800hrs for the degradation data series of Device 4	140
6.8	Time-series data of N_{eff} for the degradation data series of Device 1 .	142
6.9	Time-series data of N_{eff} for the degradation data series of Device 2 .	142
6.10	Time-series data of N_{eff} for the degradation data series of Device 3 .	143
6.11	Time-series data of N_{eff} for the degradation data series of Device 4 .	143
6.12	Time-series data of N_{iden} for the degradation data series of Device 1 .	144
6.13	Time-series data of N_{iden} for the degradation data series of Device 2 .	145
6.14	Time-series data of N_{iden} for the degradation data series of Device 3 .	145
6.15	Time-series data of N_{iden} for the degradation data series of Device 4 .	146
6.16	(a) RUL distribution based on the multinomial resampling; (b) RUL distribution based on the IMHA-replacement resampling; for the degradation data series of Device 4	147
A.1	Transfer Characteristics of (a) Device #7 and (b) Device #13 Pre- and Post-irradiation	163
A.2	Gate Leakage Current of (a) Device #5 and (b) Device #13 Pre- and Post-irradiation	164
A.3	Device-to-Device Variations of Threshold Voltage in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group	167
A.4	Device-to-Device Variations of Maximum Transconductance in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group	169
A.5	Device-to-Device Variations of Gate Leakage Current in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group	170
A.6	Evolution of Mean ΔV_{th} Values for All Devices Irradiated and Characterized at All Doses	172
A.7	Evolution of Mean Normalized $g_{m,max}$ Values for All Devices Irradiated and Characterized at All Doses	173
A.8	Evolution of Mean Normalized $I_{g,leak}$ Values for All Devices Irradiated and Characterized at All Doses	173

Chapter 1: Introduction

1.1 Problem Statement

This dissertation presents a new prognostic model for remaining useful life (RUL) prediction of semiconductor power electronic devices using the particle filter (PF) approach. The necessity of such models was outlined in the reports of the U.S. Department of Energy concerning power electronics [13], [14].

A robust prognostic model not only provides forecasting of maintenance through accurate RUL prediction, but also allows early detection and warnings of degradation during field operations. This requires a method to detect the anomalous behavior of the monitored system or component. The first important contribution of this dissertation is a novel anomaly detection technique, which is an integral part of the prognostic model. This method uses a novel singular-value-weighted distance (SVWD) in the feature space, in conjunction with the Fisher criterion between distributions of the feature space distances to detect anomalies in the failure precursor time-series degradation data of power electronic devices.

The contribution of the present work in particle filter methodology is the development and validation of a novel resampling technique, called IMHA-Replacement resampling, for the particle filter. Resampling is a critical step in many particle filters

and is required to avoid the particle degeneracy problem. An optimal resampling technique should refrain from particle degeneracy, as well as sample impoverishment, which is an issue in traditional resampling algorithms. The novel IMHA-replacement resampling was demonstrated to maintain a high level of effectiveness of particles, while maintaining a low level of sample impoverishment throughout the particle filtering process on time-series degradation data. It outperformed the state-of-the-art resampling techniques used for benchmarking, including “the classic four”: multinomial, systematic, residual, stratified resampling, as well as standalone IMHA resampling, and roughening resampling. All the techniques were implemented in the framework of a classic sampling importance resampling (SIR) particle filter.

Earlier investigations on prognostics of power electronic devices mainly focused on power devices most commonly used in industry: insulated gate bipolar transistors (IGBTs) and metal-oxide-semiconductor field-effect transistors (MOSFETs) [15]. Lacking however were investigations centered around the reliability of wide-bandgap power devices, based on GaN and related compound semiconductors. The third important contribution in the present dissertation is the first implementation of particle-filter-based RUL prediction on GaN HEMTs, which is designed for high voltage ($V_{DS} = 50\text{ V}$) and high frequency (up to 6 GHz) applications. The proposed prognostic model, first validated on IGBT modules, was also validated on high voltage high frequency (HVHF) GaN HEMTs. The failure mode of GaN HEMTs studied was voiding in the source-connected field plate induced by ohmic contact inter-diffusion under HVHF accelerated tests. The applicability of the model to devices designed to operate at high voltages and high frequency conditions, was

demonstrated.

Finally, this dissertation is the first to investigate the effect of gamma radiation on the reliability of enhancement-mode (E-mode) GaN HEMTs. Radiation is a significant driving force of degradation for GaN HEMTs in the harsh environment of Radiofrequency (RF) and microwave applications. A wide span of doses from 5 krad up to 60 Mrads (Si) was performed on commercially available E-mode GaN HEMTs. The devices were characterized pre- and post-irradiation to gauge the variations in the critical DC parameters. The wide coverage of irradiation doses facilitates observation of the complete degradation behavior of the E-mode GaN HEMTs in hard radiation applications.

In summary, the scholarly contributions of this dissertation are:

- Development of a robust particle-filter-based prognostic model integrating both anomaly detection and life prediction with a novel anomaly detection technique.
- Validation of particle-filter-based prognostics on GaN HEMTs for high voltage and high frequency applications, as well as proposed validation on radiation hardness of GaN HEMTs (discussed in Appendix A).
- Development and validation of a novel IMHA-Replacement resampling technique for the particle filter approach.

1.2 Overview of Prognostics and Health Management

To minimize the human lives and economic impact of unexpected catastrophic failures, advanced warning of potential malfunctioning and accurate predictions of the RUL of the investigated system are essential. A survey on the power electronics industry a few years ago [4] presented only 50% satisfaction rate with currently available reliability monitoring methods, showing the necessity of further research efforts in health management. State-of-the-art reliability prediction is switching from traditional techniques relying on historical field data to more advanced techniques relying on in-situ operational and environmental data. A discipline establishing the framework for these techniques is PHM [16]. The basic purpose of PHM is to detect signs of failures, provide warnings in advance, and forecast maintenance as needed. Therefore, the most important benefits of PHM includes avoiding unexpected failures and saving life-cycle cost [2].

PHM can be described as an integral cycle of five layers, as is shown in Figure 1.1. The five layers could be partitioned into three stages: observe, analyze, and act. The observe stage performs in-situ data monitoring and acquisition on the investigated system, to obtain raw data using various types of metrology equipment and sensors. In the analyze stage, the health condition of the system is assessed from the acquired raw data to determine if deviations from the normal responses and conditions has appeared. Once such deviations are detected, diagnostics are implemented to isolate the source of the issue, identify the possible failure mechanisms and evaluate the severity of damage. Based on the results of diagnostics, prognostics

is implemented to project and estimate the RUL of the degrading system. This will provide decision support to create the maintenance or replacement plan based on the actual condition of the system, i.e. condition-based maintenance (CBM) [17], [18].

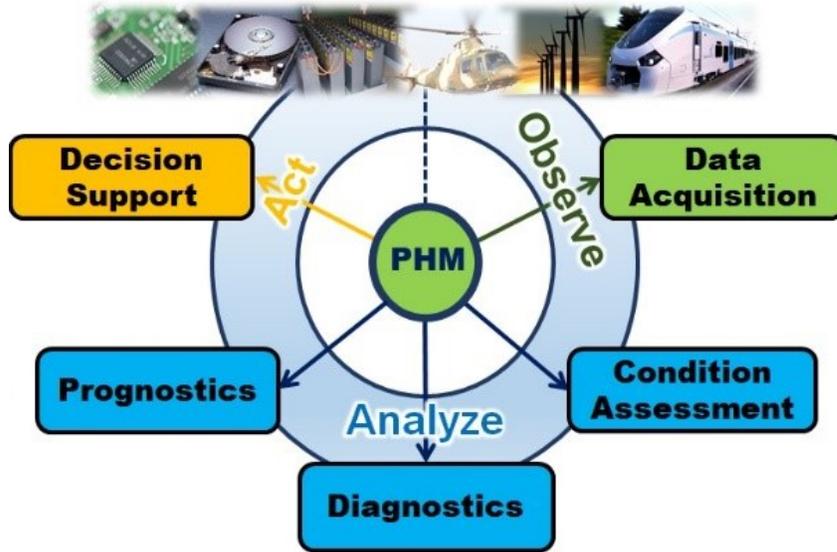


Figure 1.1: A complete PHM cycle [2]

Prognostic approaches are either physics-model-based or data-driven [16]. Model-based approaches require that explicit mathematical and physical degradation models can be derived from first principles [19]. Examples of such models for prognostics include the non-linear stochastic fatigue cracking model [20], [21] and the creep growth model based on the Norton law [22]. Based on these models, life cycle loading conditions, geometry, and material properties can be utilized to identify potential failure mechanisms and estimate the remaining useful life [23]. Given that the relevant first principles are available, model-based approaches allow estimation of damage accumulations and RUL for specific failure mechanisms, with likely much higher accuracy than data-driven approaches [24]. Data-driven approaches use sta-

tistical and probabilistic techniques based on historical information and routinely monitored data obtained from the systems to estimate the RUL [25], [26]. Generally, these approaches use statistical models that capture trends of variation from historical data for RUL predictions [27]. They assume that unless a fault shows up in the monitored system, the statistical characteristics of the monitored data remain relatively stable [16]. Faulty points, trends or patterns are detected in data collected by in situ monitoring to determine the state of health. Data-driven approaches are advantageous over model-based approaches for complex systems and where accurate modeling from first principles becomes infeasible.

1.3 Fault Detection

1.3.1 Introduction

As a critical layer in the PHM cycle, diagnostics is the process of detection, isolation of deviations from the expected behavioral patterns, which can be termed a fault, as well as identification of the cause and severity of the fault [28]. This process involves assessment of the current state of the system, and the causes of fault based on the information extracted from raw observations (measurements) [29]. Fault Detection and Isolation (FDI) techniques have been developed to provide early warnings of failure in advance and avoid catastrophic failures [30].

1.3.2 Model-based and Data-driven Fault Detection

Generally, these techniques can also be classified into model-based techniques and data-driven techniques [31]. Like the classification of prognostic approaches in the previous section, model-based techniques use a model based on first principles (mathematical models, physical laws, etc.) to describe the system and determine the current status of the system [16]. They can estimate the incipient development of faults even if faulty behavior has not explicitly shown up in the behavioral pattern of the system based on knowledge of the system architecture, the operating environment, and the governing first principles. Deviations in the behavioral patterns enable updates in the model parameters to accurately match the latest evolutions of system states. Therefore, combination of the first-principal models and the in-situ raw measurements allows model-based techniques to foresee the future evolution trajectories of key variables, and facilitate robust assessments of the health status. Model-based FDI techniques could be classified into observer-based methods, parity-space methods, and parameter-identification-based methods [32].

Data-driven techniques, on the other hand, require comparing the current behavioral patterns assessed from the latest in-situ data with the known nominal or historical data instances and operating states [33]. In FDI techniques, behavioral patterns extracted from known healthy data are often taken as a benchmark. Once deviations in behavioral patterns from the benchmark is confirmed to a high enough confidence level, a fault is considered to be detected. The deviated behavioral patterns could be further processed and compared with known faulty patterns to

isolate and identify faults, which requires a classification technique to distinguish faults occurring on different parts of the system and with different root causes [34]. Examples of common data-driven FDI techniques include Fuzzy logic, Artificial Neural Network (ANN) and Support Vector Machine (SVM).

The problem of model-based techniques is that establishing an accurate first-principle-based model would require significant amounts of expertise and efforts, especially for complex systems. In reality, assumptions have to be made from time to time to simplify the problem. Nevertheless, simplifications deteriorate the accuracy and applicability of the model to the actual situation [16]. On the other hand, the problem of data-driven techniques is the complete reliance on the historical data instances to assess the health status. In fact, misinterpretation of the system operating status could happen if the information extracted from the monitored data does not agree with either the healthy baseline or the known classified faulty patterns. Also, sensitivity to outliers may also lead to misjudgments of the status. [33] Overall, a fusion of both types of FDI techniques is often necessary for complex systems as a compromise of their respective pros and cons.

1.3.3 Supervised, Unsupervised and Semi-supervised Fault Detection

Machine learning techniques are often utilized to solve FDI problems. They can generally be categorized into supervised learning, unsupervised learning and semi-supervised learning techniques. Supervised learning is based on a fully-labeled training dataset, including nominal or healthy data and all types of anomalous data.

For all the training data, an input has a corresponding output, so the primary task of supervised learning is to learn a mapping function between inputs and outputs of training data. On the other hand, unsupervised learning does not require any labeled training data. None of inputs has a corresponding output, so inference of the underlying structure or distributions of the training data is the fundamental task of unsupervised learning [35]. Semi-supervised learning, used in the present work, stands between supervised and unsupervised learning using a mixed set of labeled and unlabeled training data [36]. It can help avoid the expertise and costs required for a fully-labeled training dataset, while improving learning accuracy considerably compared with unsupervised learning. It has been considered in the machine learning literature to reach a compromise between the availability of fully-labeled training data and the learning accuracy [37].

Recent investigations in a variety of domains have utilized semi-supervised learning techniques for anomaly detection. Ashfaq et al. [38] proposed a fuzziness based semi-supervised learning approach with the supervised learning approach of single hidden layer feed-forward neural network (SLFN) assisted by unlabeled samples to improve the classifier's performance for the intrusion detection systems in cyber security. The experimental results of intrusion detection on an evaluation dataset based on this approach showed that unlabeled samples contributed enormously to improve the classifier's performance compared to existing classifiers like naive bayes, support vector machine, etc. Zhao et al. [39] proposed a graph-based semi-supervised learning model utilizing a few labeled training data for fault detection in solar photovoltaic (PV) arrays. The model learned PV systems autonomously over

time as weather changes based on normalized voltage and current measurements, and also further identifies the possible fault type to promote system recovery. It was demonstrated experimentally that the proposed model could correctly detect and classify specific normal conditions, line–line faults, and open-circuit faults in real-working conditions. Sillito et al. [40] proposed and demonstrated an incremental semi-supervised one-class learning procedure for anomalous behavior detection in a video surveillance scenario, in which unlabeled trajectories were combined with occasional examples of normal behavior labeled by a human operator. Although the author believed that the new procedure at least as effective as existent unsupervised learning techniques, not enough demonstration of superiority using large real-world datasets was provided.

1.3.4 Anomaly Detection

Faults are deviations from expected behavioral patterns, with various natures and formats of manifestation. Anomaly detection is the process of finding patterns in data that do not conform to the expected behavioral patterns, and the data instances corresponding to the non-conforming patterns are called anomalies or outliers. Data anomalies are one important form of manifestation of faults. They provide useful and often critical information for fault isolation and identification. For anomaly detection based on semi-supervised learning techniques, labeled training data encompassing the normal behavioral patterns are used to train the model to recognize normal behavioral patterns. Applications of anomaly detection vary from fault detection

in space shuttles, aircrafts, gas turbines, etc., to fraud detection for credit cards, to cyber security monitoring [41].

There are various causes of anomalies in data, including malicious activities, system malfunctioning, or simply a random stimulus in the surrounding environment [41]. It is crucial and often difficult to identify sources of anomalies. Different anomalies have different real-life relevance. Such identification typically requires supervised techniques to perform classification to assist in recognizing different types of anomalous behavioral patterns using fully-labeled training data, which in reality could be difficult to obtain. Other difficulties encountered in anomaly detection include determination of a normal operating range covering all the possible normal behavioral patterns of the system [42], determination of a clear and precise boundary between normal and anomalous behavior [43], unpredictable variations of normal behavior in some application domains, as well as noisy data that challenges data analysis.

Data anomalies could be classified into three categories: point anomalies, contextual anomalies and collective anomalies. Point anomalies are those with an individual data instance showing anomalous behavior compared with the rest of data. This is the simplest type of anomalies. Contextual anomalies are data instances that exhibits anomalous behavior only in a particular context. The anomalous behavior is identified based on the expected values for the behavioral attributes within a specific context [41]. Collective anomalies are a collection of related data instances showing anomalous behavior with respect to the entire data set. A data instance in a collective anomaly may not seem anomalous when viewed individually, but their

occurrence together as a collection shows anomalousness.

1.4 Dynamic State Estimation and Particle Filters

1.4.1 Bayesian Filters for Dynamic State Estimation

To provide warnings of failure in advance, model-based prognostic approaches require future state estimation of variables indicating the health status of the system based on first-principle models. The evolution of these variables over time could be tracked through a sequence of noisy measurements taken at discrete time steps and modeled through difference equations [44]. In this sense, prognostics is essentially a dynamic state estimation process based on sequential modeling at discrete time steps. Dynamic state estimation tracks and makes inference about a dynamic system, based on a system process model and a measurement model. A system process model in state estimation for prognostics describes the sequential evolution of the state variables over time, to help identify any degradation of the system and make prediction of future states and time-to-failure. A measurement model, on the other hand, describes the noises of raw measurements and restores expected actual values of measurement. In other words, a system process model makes inferences of the future states while tracking the actual measurements, while the measurement model outputs measurements of the state variables that update the inference results. This essentially becomes a recursive Bayesian procedure in dynamic state estimation, in which the state inference results (prior) is updated with the latest measurements (likelihood) to obtain the updated estimation (posterior). The current state estimate

is only based on the most recent state available due to the condition of Markov property that given the current state, the future state is independent of all the previous states of the system. A recursive Bayesian filter consisting of an inference step and an update step is formed following this procedure for solving dynamic state estimation problems. Among the common Bayesian filters, particle filters are capable to handle the problems with non-linear process models and non-Gaussian state estimation probability density functions (PDFs) compared with standard/extended Kalman filters [44].

1.4.2 Particle Filters

Particle filtering (PF) is a state-of-the-art model-based Bayesian inference approach solving nonlinear and non-Gaussian state estimation problems. It is based on sequential Monte Carlo simulations utilizing Bayes theorem and Markov property [45]. The basic idea of a PF is to represent the posterior PDF with a large number of particles and their associated weights, which are updated iteratively at every time step. In this sense, the posterior PDF representation is not restricted to Gaussian distributions. Another significant advantage of the sample-based representation of PFs is that it can model non-linear evolutions of state variables without relying on any local linearization technique or coarse functional approximation [46]. PFs have been used in various sectors of field applications including robotics, automation, computer vision artificial intelligence and even chemical engineering.

A PF performs dynamic state estimation following the recursive inference and

update procedure described earlier. The current state is inferred using the posterior PDF of the previous state, obtaining a prior PDF. Then the latest measurements at the current step are used to update the prior PDF, obtaining a posterior PDF using Bayes theorem. In the PF approach, both prior and posterior PDFs are represented by a set of particles sampled from the distributions with associated weights denoting discrete probability masses. The particles are initiated at the beginning of state estimation processes and their positions and weights are recursively updated using the system model, measurement model and the actual measurements [47].

1.4.3 Literature Review of Particle-Filter-Based Prognostics

In prognostics, RUL predictions are essentially long-term predictions of future states made at a specific point upon a triggering event such as detection of a fault, or a designated checkpoint is reached. For particle filters, this is realized through repetitive propagation of the particles without updating their weights at each time step. The propagation continues until a specified number of particles reach the predefined failure threshold. Many previous studies have reported PF-based prognostics of electronic devices and components, including IGBTs, power MOSFETs and electrolytic capacitors. The following is a detailed review of the past investigations on PF-based prognostics, focusing mainly on semiconductor power electronic devices.

Orchard et al. [29] presented an online PF-based model framework for fault diagnosis and failure prognosis in the blades of a turbine engine. The model framework

consisted of two autonomous modules, a fault detection and identification (FDI) module using a nonlinear dynamic state model and a PF algorithm to calculate the probability of fault condition (crack appearance). A feature called the tangential blade position (TBP) was generated to map the crack length in the turbine blades. The state PDF estimates were also computed as initial conditions in the prognosis module. The failure prognosis module calculated the p-step ahead long-term prediction and estimated the RUL PDF of the faulty turbine blades using a PF-based algorithm based on a nonlinear state-space model with unknown time-dependent parameters. The prognosis module could predict the evolution of the PDF of the crack length. The proposal of this model framework was novel and ground-breaking, especially for the implementation of PF in both diagnosis and prognosis. However, the reliance of diagnosis on just one feature may be questionable. Also, it was unclear how the nonlinear mathematical model in the FDI module was derived.

Celeya et al. [48] implemented and compared three different approaches for prognostics of power MOSFETs including data-driven Gaussian process regression (GPR), and model-based extended Kalman filter and particle filter. Die-attach degradation under thermal overstress was identified as the primary failure mechanism, and on-state resistance $R_{DS(ON)}$ was selected as the precursor. In the aging experiments, the increase in junction temperature was observed with the increase of $R_{DS(ON)}$ showing the deterioration of thermal dissipation as the die attach degrades. An exponential degradation model was proposed for power MOSFETs. The results of accelerated aging tests for six power MOSFETs were used for prognostics. The RUL predictions were made at a number of different time points. The α - λ performance

metric was used to evaluate the RUL prediction results. It turned out that the GPR approach could only make predictions at a much later time compared with model-based Bayesian filters, due to the lack of degradation model incorporation. Between the two Bayesian filters, the particle filter provided steadier and overall more accurate prediction results than the extend Kalman filter, although their results were mostly close. However, no fault detection technique was utilized to make a methodic selection of the RUL prediction triggering point.

Saha et al. [49] implemented the PF method on the prognostics of IGBTs in avionics systems. The devices were aged under thermal overstress by turning the gate on/off to keep the temperature cycling between 270°C and 305°C, until the occurrence of thermal runaway and latch-up failures. The shifts of the tail collector-emitter current I_{CE} while turning off the devices, were used as the failure precursor. The prediction algorithm conducted regression of the tail I_{CE} using an exponential degradation model to obtain the polynomial model parameters of each cycle. These parameters were fed into the PF-based degradation model to form the state vector. The system importance resampling algorithm was adopted on particles to make RUL predictions. The author proposed a framework of extracting features from raw data and using them to monitor the system behavior by learning the model parameters. As soon as a diagnostic trigger shows up, the RUL estimation is triggered. However, the diagnostic trigger was not specifically addressed and the RUL estimation was chosen to be made at an arbitrary time.

Patil et al. [50] developed a PF-based prognostic approach that could detect anomalies and predict the RUL of non-punch through (NPT) and field stop (FS)

IGBTs. The devices underwent power cycling stresses until failure. X-ray analysis before and after the power cycling tests showed the failure mode of die attach degradation. The on-state collector–emitter voltage $V_{CE(ON)}$ and the on-state collector–emitter current $I_{CE(ON)}$ were used as failure precursors and were monitored in-situ throughout the test. The failure threshold was defined as a 20% increase in $V_{CE(ON)}$. The anomaly detection was performed using Mahalanobis distance (MD) computed from $V_{CE(ON)}$ and $I_{CE(ON)}$ parameters. Once an anomaly was detected, the RUL prediction was triggered using the PF algorithm. The system model for PF implementation was obtained using a least squares regression of the $V_{CE(ON)}$ data. The mean time to failure (MTTF) estimates of the RUL revealed an error of around 20% at the time of anomaly detection. The contribution of this work was MD-based anomaly detection. However, the prediction error of 20% was significant. A better system model was needed to reduce the error.

Haque et al. [51] proposed an approach for the RUL estimation based on auxiliary particle filter (APF). APF was supposed to reduce sample impoverishment while maintaining the diversity in samples by introducing the index of particles in the previous step as an additional variable for resampling. The simple slope-based method was used to identify and divide the entire degradation curve into three regions: healthy region, constant increase region, and exponential increase region. For each region, different state transition equations were applied considering the different variation patterns and degradation severity of $V_{CE(ON)}$. The proposed method was shown to reduce estimation variance through introduction of the sample index in resampling. Power cycling tests were conducted on seven IGBT modules

and the observed failure mode was wire bond life-off. A 20% increase in $V_{CE(ON)}$ was used as a failure threshold. The results of RUL estimation on the obtained IGBT degradation curves were compared between APF and SIR PF algorithms based on a variety of particle numbers and different regions, and APF was shown to outperform SIR PF overall in root mean square error (RMSE). While using APF for the RUL estimation of IGBTs was a solid contribution, dividing the degradation curve into three different regions seems too generic and whether it can be applied universally to all the IGBT degradation curves under power cycles is questionable. Also, there was no diagnosis process and no point was specified at which the RUL estimation was initiated.

Rigamonti et al. [52] developed a particle filter-based prognostic model for the RUL estimation of aluminum electrolytic capacitors in electrical automotive drives. The author focused on gradual failures of capacitors in which equivalent series resistance (ESR) was commonly used as a degradation indicator. The failure mechanism of the capacitor was considered to be the vaporization of electrolyte, which is strongly affected by the capacitor operating conditions, especially the working temperature. The vaporization led to an increase in ESR, and when its value doubles based on the initial value, the capacitor was considered failed. To take into account the influence of temperature, the author introduced a novel degradation indicator, the ratio between the ESR of the degraded capacitor and the ESR value on a new capacitor at the same temperature. This ratio was independent from the measurement temperature. A standard prognostic model was used on both simulated and experimental degradation test results for the RUL estimation. The author

first performed measurements to obtain the parameters of the relationship between ESR and working temperature for new capacitors. Then a simulation of realistic capacitor degradation was conducted, and the prognostic model was implemented. The performance of the RUL estimation was evaluated with five metrics: precision, accuracy, steadiness, coverage, and risk level. The accelerated degradation tests were performed on commercial capacitors. The temperature was kept constant at 418 K, and due to limited test time, a threshold of 130% ESR was used, and the PF algorithm was implemented in the same way. The novelty of the work was on a new degradation indicator and the first PF implementation on capacitors. Nevertheless, no anomaly detection technique was used and the test suffered from lack of sufficient measurements (only seven measurements were available).

Wu et al. [53] presented an improved particle filter (IPF) method for the RUL estimation of MOSFETs. The on-resistance R_{on} was used as the degradation indicator. The approach first used strong tracking Kalman filter (STKF) as the importance function to update the particles in the sampling process with the latest observation. At the resampling step, the Metropolis–Hastings algorithm was used to replace the regular resampling algorithm. The proposed approach was validated on two data sets in the “MOSFET thermal overstress aging data set”. A failure threshold of 0.045Ω increase in R_{on} was defined, and several different points was chosen for the RUL observation. The results showed lower RMSE for the improved particle filter algorithm than that of the conventional particle filter. Although the contributions of the work seemed clear, there was not a clear justification of how the STKF worked and why it was superior to the conventional importance function.

Also, the author acknowledged influence of linearization by introducing STKF into the PF algorithm.

Chen et al. [54], [55] focused on the selection of dynamic time features as health precursors for prognostics of power MOSFETs in DC-DC converters. The author reviewed the common MOSFET failure modes and the degradation process, in which time-dependent dielectric breakdown (TDDB) was determined to be the failure mode studied in the work. When it came to the selection of healthy precursors, besides the accepted precursors including threshold voltage, on-resistance and transconductance that could indicate the degradation of MOSFETs, the turn-on time was identified as a new online health precursor of MOSFET. It was derived that as gate oxide traps accumulated, the equivalent MOS capacitance C_{gd} would increase, thereby increasing the gate charging and discharging time and therefore the turn-on time when power MOSFET was applied as a switch. The accelerated degradation test (ADT) was used to obtain the evolution curves of the four health precursors, in which the stress type was a high gate bias (gate-source voltage) of 53 V at a constant 25 °C temperature. To reduce the effect of thermal stress, no drain-source bias was applied. A RUL estimation based on the particle filter was performed on the turn-on time degradation curve. The RUL prediction was triggered at 960 minutes. Since significant distortion in the turn-on time evolution over 116 ns was observed, the failure threshold was set to be 116 ns. While identifying the turn-on time as a new failure precursor for MOSFETs under TDDB failure mode was a contribution, the implementation of PF in the RUL estimation was coarse. The prediction point was selected randomly. The biggest problem is that the turn-on time evolution before

116 ns seems linear, which means even linear extrapolation may be enough to make accurate prediction in this case. Also, the author did not specify the process model used.

Kwon et al. [56] utilized a model-based approach to detect and track the initiation and propagation of cracking in interconnects and predict the time-to-failure of interconnects using particle filtering. RF impedance had been found to be sensitive to incipient changes in interconnects and could provide early warnings of interconnect failures. Therefore, the author used impedance analysis, in which the time domain reflectometry (TDR) reflection coefficient was selected to measure RF impedance. An analytical model proposed in the previous work to simulate the progression of TDR coefficient with crack size was used, in which for cracks under 1 mm, the relationship between TDR coefficient and crack size was assumed to be linear. The Paris' Law was used to model the fatigue crack growth. Mechanical fatigue tests were conducted applying a cyclic shear stress condition on solder interconnects with in-situ monitoring of RF impedance. The test results confirmed consistent increase and advanced warning of the RF impedance with the degradation of solder joints. Particle filtering approach was used to predict the time-to-failure of the solder interconnects when the TDR reflection coefficient increased by 5%. This study presented the impedance analysis method of monitoring the fatigue-induced cracking in interconnects. However, the prediction point was also chosen empirically, and there was also an assumption that the crack size would always remain under 1 mm, in order for the assumption of linear relationship between the TDR coefficient and the crack size to hold. However, this may not always be true in reality.

It can be concluded from the above review that most of the available works on PF-based prognostics of electronics selected the triggering point of the RUL prediction empirically. No methodical approach for fault detection and identification was implemented. PHM is not complete without a diagnostic process identifying faults before triggering the RUL prediction. Although the framework of a comprehensive prognostic model integrating both anomaly detection and RUL prediction of electronics has been proposed, as is in Orchard et al. [29] and Patil et al. [50], they either were unclear in the fault detection model or lacked a robust system process model that led to large prediction errors. The present investigation contributes by introducing a novel anomaly detection technique and a system process model integrating both the test loading conditions and the physics of failure related to the failure mechanism.

1.5 Power Electronics and IGBTs

1.5.1 Power Converters and Power Electronics

Power electronics concerns the processing of electrical power using electronic devices, the key of which is a switching converter [57]. The applications of power converters in industry include aerospace, automotive, marine equipment, utility power supplies, etc. One of the most reliability-demanding applications of power converters is offshore wind turbines (OWTs). Figure 1.2 shows the topology of a doubly-fed induction generators (DFIG) system, which is the most common configuration in wind turbine power converters [3]. This configuration includes a generator-side and

a grid-side converter with a DC-link in connection. Each side converter consists of stacks of half-bridge power electronic modules (usually IGBT modules) connected in parallel, as is shown in Figure 1.3. The semiconductor power electronic devices (transistors, thyristors, etc.) inside these modules, performing power conversion, are the kernel components of power converters.

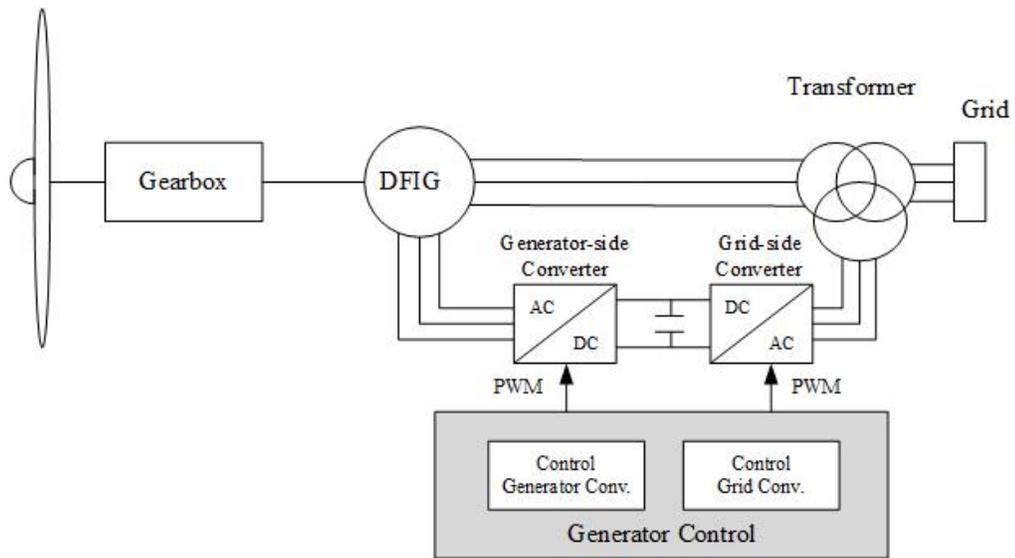


Figure 1.2: Generator topology and control of a DFIG system [3]

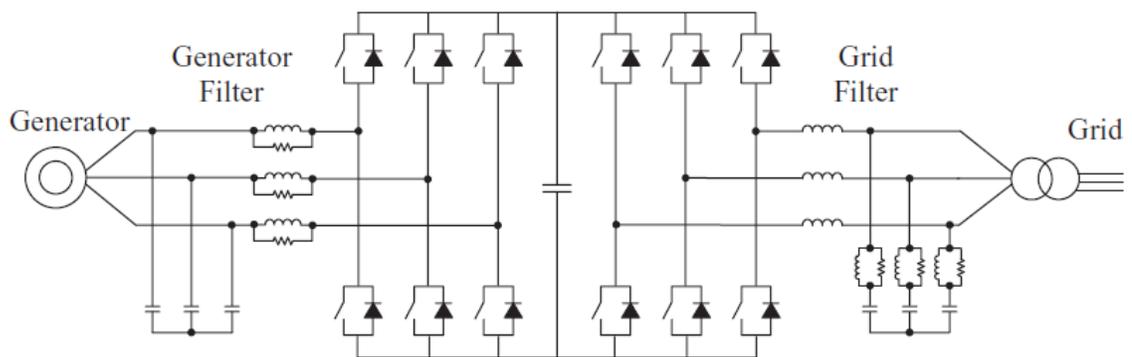


Figure 1.3: Back-to-back converter of a wind turbine with paralleled half-bridges in each phase module to provide the required current capacity [4]

Many of the power converter applications are safety-critical because of the catastrophic consequences of failure, and the zero-defect concept has been proposed placing rigorous reliability requirements on the entire system [58]. Power converters, particularly power electronic devices, are of great reliability concerns in the field operation. Various types of stressors, such as temperature cycling, mechanical vibration, and humidity, could impose stresses on the power devices in field applications [59]. A general survey based on over 200 products from 80 companies [60] showed that the semiconductor die and soldering failures in power electronic device modules were responsible for 34% of converter system failures. A reliability survey on OWTs [61] showed that the power converter assembly reported the second highest failure rate and downtime among all the subsystems.

1.5.2 IGBT Devices and Modules

An IGBT is a three-terminal semiconductor power device with four alternating layers (P-N-P-N), widely used as power electronic switching devices in systems requiring medium-to-high power (10 kW-3 MW) and medium frequency (20-200 kHz) [62]. It was first experimentally demonstrated in 1979 by Baliga [63]. The first commercially-available IGBT device came from General Electric in 1983 [64]. Common applications of IGBTs include power converters, electric vehicles, locomotives, refrigerators, air-conditioners, etc. IGBT combines the gate-driving characteristics of MOSFETs with the high-current and low-saturation-voltage attributes of bipolar junction transistors (BJTs) in a single device. It is a monolithic integration of power MOSFETs and BJT

devices combining the best characteristics of both devices to achieve optimal device characteristics [62]. The main advantages of IGBTs include low on-state voltage drop with high on-state current density, low gate-driving power with simple gate-driving circuit thanks to the metal-oxide-semiconductor (MOS) gate structure, as well as wide safe operating area (SOA) [65]. Due to these significant advantages, IGBT has been the most popular type of power electronic devices in industry. Power converters, vehicles, refrigerators, air-conditioners, etc. are all among common applications of IGBTs. A survey on industrial power electronics showed that IGBT took up 42% of all the power devices in power converters in the industry [15].

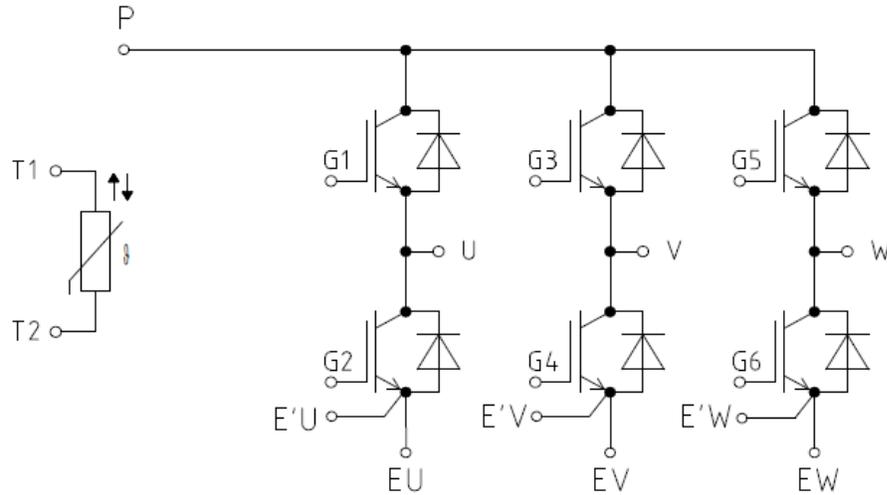


Figure 1.4: Circuit diagram of Infineon FS20R06W1E3 IGBT module [5]

As was mentioned earlier, power converters for high power applications consist of stacks of half-bridge or single-switch modules, to meet the required power level. In wind turbine applications, for example, 4-fold generator-side and the 3-fold grid-side modules are commonly used [6]. Figure 1.4 shows the circuit diagram of an Infineon FS20R06W1E3 IGBT module. In the diagram, G_1, G_2, \dots, G_6 denote the gates

of the IGBTs on the six pieces of die. U , V , W denote the AC output pins of the three phases of AC. EU , EV , EW denote the emitter output pins, while $E\cdot U$, $E\cdot V$, $E\cdot W$ denote the spare emitter output pins of the three phases, respectively. P is the collector pin shared by the six IGBT dies. $T1$ and $T2$ are the pins of a negative temperature coefficient (NTC) thermistor, which is integrated as a temperature sensor inside the module to enable accurate measurement of chip temperature.

1.5.3 IGBT Module Package

The package of power electronics refers to the package of multiple chips of power electronic devices and their interconnections for signal, power transmission and heat dissipation [66]. The package also provides electromagnetic interference (EMI) shielding, electrical conducting/insulating, housing protection from environmental contaminants and structure support. The package components are the supporting components of power device chips. For IGBTs, there are generally two different packaging types: press-pack technology and module technology. Press-pack technology is characterized by high reliability due to the elimination of bond wires, and double-sided cooling [67]. However, limited power handling capability as well as costs hindered it from widespread applications. Module technology, illustrated in Figure 1.5, remains the common packaging technology in commercial devices. Herein, the dies (IGBTs, diodes) are electrically connected by aluminum or copper bond wires. The dies are soldered to an insulating ceramic substrate called direct copper bonded (DCB) substrate. The substrate is soldered to a copper base plate, which

is in turn attached to the heat sink with a layer of thermal grease. The presence of bond wires and solders introduces a significant source of reliability issues to the package of IGBT modules, which is discussed next.

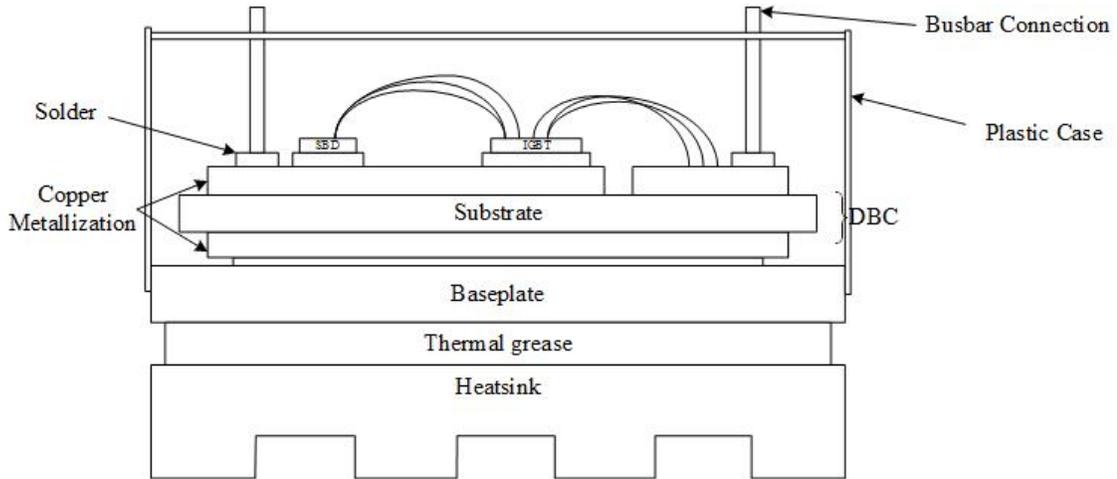


Figure 1.5: Module Package Structure for IGBTs [5]

1.5.4 Package-related Failure Mechanisms

Failure mechanisms of power electronics can generally be categorized into two types: on-chip failure mechanisms and package failure mechanisms. On-chip mechanisms are those occurring on active devices and on-chip interconnections, such as electrical static discharge (ESD) and electromigration. These mechanisms are associated with the power device chips. Package failure mechanisms, on the other hand, are the more frequent type encountered in power electronics [6]. There are a number of stressors resulting in package failures, including temperature, humidity, vibration, contamination, etc. However, according to a study of different stressors in power electronic systems [68], the temperature cycling and the steady-state

temperature are the most critical stressors influencing the failure of power devices. For example, the electrical traction drive for an urban tram may experience 10^6 – 10^8 power cycles, with temperature swings up to 80°C , during its lifetime [69]. The criticality of temperature cycling results from coefficients of thermal expansion (CTE) mismatches between adjacent package materials [59]. Both on-chip and package failure mechanisms can destroy the devices and the ones of IGBTs have been extensively investigated and well understood. As the more common type of mechanisms, package failure mechanisms of IGBT modules are of greater concerns and was selected for prognostic modeling in the present dissertation.

As was discussed in the previous section, package components can be regarded as ‘accessories’ of IGBT chips. In long-term power/temperature cycling conditions, components such as bond wires and solder interconnects lead to several thermo-mechanical reliability issues. The module package consists of stacking layers of different materials with different CTEs. The combination of CTE mismatches and temperature swings leads to stress concentration near the interfaces of different layers or in nearby weak spots. Consequently, voids and microcracks start to grow and expand, and eventually, partial or even complete detachment of bond wires may occur. Hence, significant degradation or interruption current and heat dissipation would occur, resulting in malfunctioning of the IGBT module or even catastrophic failures. The most common failure modes under such circumstances include bond wire lift-off, bond wire heel cracking and solder joint fatigue.

The fatigue of bond wires has been discussed in earlier studies as a common failure mechanism in power modules [70], [71]. Many of the bond wires in power

modules are bonded onto the active area of power devices (IGBT and freewheeling diodes), which is the heat source in the device operation. The high magnitude of temperature swings and the self-heating of wires due to current flow, combined with the significant CTE mismatch between bond wire metals and Si, result in high susceptibility of bond wires to fatigue damage. Typically, cracks grow near the bond wire/chip interface, leading to bond-wire lift-off.

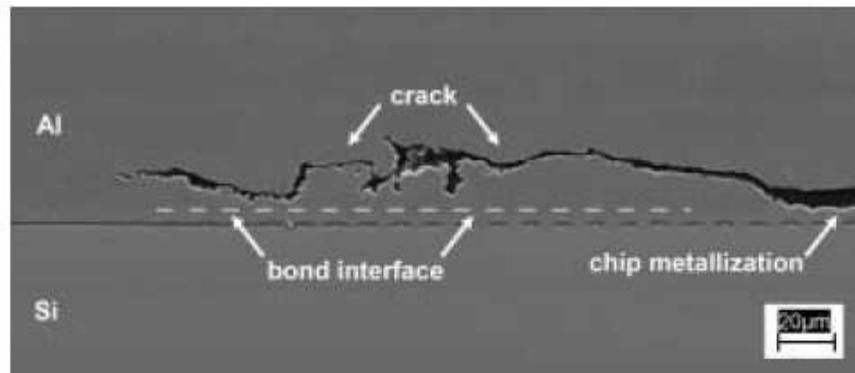


Figure 1.6: Crack growing closely above the bond wire/chip interface, leading to bond wire lift-off [6]

Another common failure mechanism of IGBT modules is fatigue of solder interconnects, which is also associated with the thermo-mechanical stress caused by temperature swings. The most susceptible components of this failure mechanism are the solder joints between ceramic substrates and copper base plate [72]. The crack typically starts from the edges of solder joints and propagates towards the center portion, finally leading to total delaminations. The consequence of solder interconnect delamination is the deterioration of heat dissipation capability and risks of device failure due to overheating. Overheating may also accelerate the occurrence

of bond wire failures discussed earlier. The significant mismatches in the CTEs between ceramics and copper, again combined with high temperature swings since the solder joints lie on the major pathway of thermal flow, contribute to the criticality of this failure mechanism.

1.5.5 Accelerated Life Testing and Power Cycling Test

In reliability evaluations, it is often necessary to acquire reliability information in a much shorter time frame than the anticipated time-to-failure of the studied component or system under normal operating conditions. Consequently, reliability assessment methods that shorten the test time while exhibiting the same wear-out degradation and failure process, such as accelerated life testing (ALT), emerge and gain widespread popularity. ALT allows reliability models to be structured for the most relevant type of stresses in a controlled manner across a wide range of stress levels [73]. Accelerated testing (AT) methods could be divided into qualitative and quantitative methods. Quantitative ALT stresses the component or system to degrade and fail in a same manner as the normal use conditions, to estimate the lifetime distribution [74].

For power devices, it has been discussed previously that power cycling critically affects the reliability of the package. In this sense, validation tests of high temperature and temperature cycling have been a key driver in improving the reliability of power modules. The accelerated power cycling test is an important test to assess the reliability of power modules under the temperature cycling stressor [75], [76]. It runs

the device under tests (DUTs) under power cycles with short cycle periods and high temperature swings to drastically accelerate package component degradations and intrinsic failures.

The establishment of a power cycling test apparatus can be referred in the relevant Joint Electron Device Engineering Council (JEDEC) standard [77]. In power cycling tests, the DUT are generally run under load pulses with constant DC input power. At the on state, the DUTs are actively heated by the conduction losses and the temperature of the DUT increases rapidly. As soon as the temperature rises to the desired maximum value, the applied input power is paused and the temperature decreases under the influence of an external cooling system, until the desired minimum value is reached. The heating and cooling time combined is defined as a cycle period and it is repeated until the DUT fails. During each period, data collection is performed at points of interest. To obtain the desired junction temperature swing ΔT_j and mean junction temperature $T_{j,mean}$, the duration and amplitude of load pulses can be adjusted accordingly [78].

1.6 HF GaN HEMTs

1.6.1 GaN Devices

Despite the current popularity of Si power devices including IGBTs, with miniaturization and high power requirements, the fundamental limits of Si technology have become more and more pronounced [79]. Table 1.1 shows the physical properties of traditional semiconductor materials including Si, GaAs and some

Table 1.1: Physical properties of important semiconductors for power devices [1]

Material	Bandgap E_g (eV)	Electron/Hole Mobility ($cm^2/V \cdot s$)	Maximum Electric Field E_c (MV/cm)	Thermal Conductivity $k_{thermal}$ (W/cm · K)
Si	1.1	1350 (n)	0.3	1.5
GaAs	1.4	8500 (n)	0.4	0.5
SiC	3.26	900 (n)	2.0	4.5
GaN	3.39	1000 (n)	3.3	1.3
Diamond	5.45	3800 (n)/4500 (p)	2.0	4.5

wide-bandgap semiconductor materials, including silicon carbide (SiC), GaN and diamond. Wider bandgaps allow materials to operate at higher temperature, making them attractive for high power applications [80]. Moreover, wide-bandgap materials show significantly higher maximum electric field, which makes them preferable in high-voltage applications because it allows the devices to be biased at higher drain voltages [81]. Diamond has the best theoretical properties, but the high cost and the underdeveloped device technology have prevented it from commercialization in the near future.

GaN-based devices, on the other hand, have emerged as a strong candidate for the next generation high-efficiency semiconductor electronic switching devices due to the inherent advantages of the GaN material, as well as the piezoelectric and spontaneous polarization effects in AlGaN/GaN heterojunction that result in 2-dimensional electron gas (2DEG) channel with electron mobility in excess of $2000 \text{ cm}^2/V \cdot s$ [82]. This is much higher than the electron mobility of Si, accounting

for the higher frequency and the superior performance of the devices. In addition, high-electron saturation velocity of around 3×10^7 cm/s and excellent potential of thermal conductivity are also desirable properties [83].

1.6.2 HEMTs

A High-electron-mobility transistor (HEMT), also known as heterostructure FET (HFET) or modulation-doped FET (MODFET), is a field-effect transistor incorporating a junction between two materials with different bandgaps, named heterojunction [84]. In a MOSFET or metal–semiconductor field-effect transistor (MESFET), the channel is contained within the n-type channel that forms the Schottky barrier and is a doped region. However, in a HEMT, the channel lies in an undoped layer near the heterojunction of two different semiconductor materials [85]. Combining the inherent advantages of wide-bandgap semiconductor materials, and the high electron mobility, high carrier density of 2DEG in the Aluminium Gallium Nitride (AlGa_N)/Ga_N heterostructure, Ga_N HEMTs have shown significant promise as the next high-power and high-frequency power electronic devices.

The polarization effect of the AlGa_N layer on the Ga_N layer results in positive polarization charges at the AlGa_N/Ga_N interface and negative polarization charges at the top of the AlGa_N layer, leading to formation of an electric field in the AlGa_N layer [7]. Since the AlGa_N layer is n-doped, and the fermi-level of Ga_N is lower than that of AlGa_N, electrons conducting in AlGa_N move across the heterojunction into Ga_N barrier layer, closing up the fermi-level difference until the Fermi level becomes

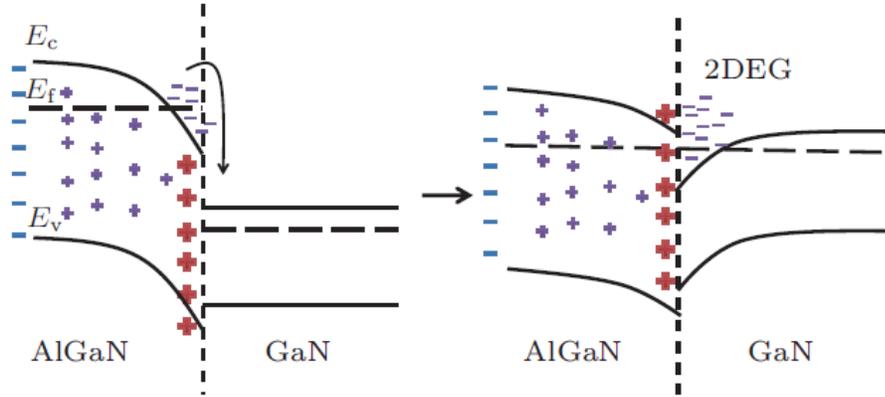


Figure 1.7: The Schematic of an AlGaN/GaN heterojunction showing the formation of 2DEG layer [7]

flat again. The resultant bending of conduction band in the GaN layer near the interface and the discontinuity of conduction band at the interface form a potential well confining free electrons. The electrons in the well are only free to move around in the plane in parallel with the heterojunction, so a very high sheet carrier density can be achieved. Since there is no doping in the channel layer, there is no ionized impurity scattering, which contributes to the high mobility of the channel layer.

1.6.3 High Frequency Operation of GaN HEMTs

The "high frequency (HF)" regime referred in the present dissertation (up to 6 GHz) constitutes part of the electromagnetic spectrum with frequency in the range from 20 kHz to 300 GHz [86], which is typically referred in the literature as "radiofrequency (RF)". The pulsed HF operation of GaN HEMTs is realized by switching the gate bias of the transistor on and completely off to ensure that the device would go through full transients for every pulse [87]. The fixture generally

consists of a DC pulse supply and a GaN HEMT HF operation circuit. The pulse supply outputs the gate and drain bias, and it controls precisely the alternation of on and off-state voltage. The HF operation circuit consists of the GaN HEMT, HF input and output matching units, gate and drain bias transmission lines and other associated components.

The high-power, high-frequency and high-temperature capability makes GaN HEMTs an outstanding candidate for radio wave and microwave applications like satellite communication, radar, etc., which often require working reliably in harsh aerospace, military and nuclear environments. GaN HEMTs have demonstrated power densities of 6-9 W/mm, nearly one order of magnitude higher than those of Si devices and GaAs HEMTs [88]. Also, nowadays commercial GaN HEMTs with drain voltages of 50 V are available, thanks to the use of field plates that reduces the peak electric field in devices [87].

1.6.4 Failure Mechanisms of HVHF GaN HEMTs

In terms of reliability, superior lifetime has been reported from ALTs on GaN HEMTs, and multiple failure mechanisms have been reported as well [89]. For example, a common issue associated with HF operations of GaN HEMTs is the drain current reduction and gain compression, which is considered to result from surface state traps. This problem can be relieved from SiN passivation [90]. Although largely reduced, surface traps still exist in these devices. In fact, a common root cause of GaN HEMT degradation is defect generation in the AlGaIn layer under high vertical

electric field near the gate. These defects trap electrons, reducing the sheet carrier concentration and the drain current [91].

Further investigations and characterizations of the failure mechanisms under HF operations have been conducted. Chen et al. [92] studied the catastrophic failure mechanisms limiting the survivability of a T-gate GaN HEMT under HF overdrive. Simulations were performed using a 4-finger $200\ \mu\text{m}$ GaN HEMT device model, with two catastrophic failure mechanisms identified: sudden failure due to the forward turn-on exceeding the burn-out limit at low quiescent drain-source voltages ($< 10\ \text{V}$) and gate-drain reverse breakdown at higher quiescent drain-source voltage due to high peak drain-gate voltage. For wear-out mechanisms, Coffie et al. [93] investigated the temperature and drain-to-source voltage dependency of the lifetime of GaN HEMTs under HF stresses. It was discovered that the lifetime decreased with increasing V_{DS} but increased with increasing temperature. The degradation process was not diffusion limited, and hot-carrier-induced degradation was believed to be the dominant mechanism. Valizadeh et al. [94] also believed that hot carriers trapped in the surface or barrier layer led to the degradation based on DC and HF stress tests, in which no evident differences in degradation mechanisms between DC and HF stresses were found.

In more recent papers, Sasikumar et al. [95] conducted HF-ALTs on low gate leakage operational S-band GaN HEMTs, and discovered a 1.2 dB degradation in HF output power. The degradation was attributed to an increase of on-state resistance resulted from an increase in the concentration of various deep states between E_C of 1.6 and 3.0 eV. A later paper from the same group related the degradation of

GaN HEMTs under HF-ALTs to defects with $E_C = 0.57eV$ [96]. Gajewski et al. [69] reported the results of the DC-ALT on 28 V and 40 V GaN HEMTs monolithic microwave integrated circuits (MMICs) developed by Cree, with the observed failure mechanism to be ohmic contact inter-diffusion. The degradation results of 50 V devices under HF-ALT was also reported, with the degradation and failure mechanism of voiding in the field plate also due to ohmic contact inter-diffusion in the source-connected field plate. The introduction of field plate structure in the vicinity of gate helped reduce the peak electric field in the channel caused by gate contact at the drain side of the gate edge [97], improving the power performance and breakdown voltage of devices. Nevertheless, source-connected field plates have been discovered to degrade with void growth and coalescence inside caused by ohmic contact inter-diffusion between the field plate and the source contact under HF-ALTs. Consequently, there was a gradual loss of function for the field plates, exacerbating defect generation in the active layers and degradation of power output.

In summary, there has been no clear consensus on the exact degradation and failure mechanism that leads to HF output power degradations. More investigations on the mechanisms under HF-ALT are still needed to enhance the understanding. In the present work, the failure mechanism adopted was the actual mechanism observed and reported corresponding to the degradation data used for prognostic modeling.

1.6.5 HF-ALT

Currently, many GaN device manufacturers have implemented JEDEC and Automotive Electronics Council (AEC) qualification on their products to qualify the developed devices, eliminate early failure modes and enhance their understanding of the device reliability. The most common tests include high temperature reverse bias (HTRB), high temperature gate bias (HTGB), high temperature high humidity reverse bias (H3TRB), etc. As qualification tests, they are typically run for only 1,000 hours. Intrinsic reliability tests, like DC and HF ALTs, have also been developed and conducted to discover the wear-out failure mechanisms of GaN HEMTs. Similar to the power cycling tests of IGBT modules, HF-ALT is an intrinsic reliability test that runs until parametric failure or catastrophic failure or end of the test. HF-ALTs are usually performed using a commercial HF reliability test system under HF drive conditions, with each DUT individually mounted and soldered to a metal carrier, which is in turn attached to a test fixture. Each test fixture has an independent heater block to control and maintain its temperature constant, in order to achieve a constant target junction temperature throughout the test. There are also interfaces for DC and HF stimulus and HF input/output matching units [98]. A semiconductor parameter analyzer (SPA) is usually embedded in the system to supply DC signals and perform DC characterization before and after the test. Computer controls the on/off state of the heater block, performs data acquisition and commands the SPA for in-situ device characterization. The DUTs are stressed at the nominal drain operating voltage and the HF input unit supplies a high-frequency continuous wave

(CW) signal. The selection of HF input drive level is such that a high HF compression level of around 3-6 dB and the desired junction temperature can be achieved, which can be encountered in some field applications [69].

1.7 Dissertation Scope and Outline

In this dissertation, a comprehensive prognostic model, which integrates anomaly detection and RUL prediction, is developed for lifetime estimation of power electronic devices. Unsupervised machine learning techniques are utilized for baseline establishment, and then semi-supervised learning is used for anomaly detection. The Particle filter approach is used for RUL estimation bring in stochasticity for the prediction. A novel resampling technique is developed that outperforms the state-of-the-art resampling techniques. The developed model is validated on time-series degradation data of IGBT modules with fatigue-related failure mechanisms under power cycles, as well as degradation data of GaN HEMTs with ohmic contact inter-diffusion as the failure mechanism. The appendix of this dissertation addresses the stability of DC parameters of enhancement-mode GaN HEMTs under gamma irradiation. Commercial E-mode GaN HEMTs were irradiated with a wide span of total doses from 5 krad up to 60 Mrads, and DC parameter degradations are reported. The degradation evolutions of the critical DC parameters will enable another potential application scheme of the prognostic model.

Chapter 2 introduces the idea and formulations of non-linear Bayesian tracking problem, as well as the particle filter approach in general. The sequential importance

resampling (SIR) particle filter used in the present work is introduced thereafter.

Chapter 3 presents in detail the introduction of the prognostic model. The general model framework is discussed first, and then the anomaly detection technique is introduced, including feature construction, feature extraction based on principal component analysis (PCA) and k-means clustering, the novel distance measure of singular-value-weighted distance (SVWD) in the feature space, and the outlier probabilities determination based on Fisher criterion. The development of the system process model and the measurement model is then discussed.

Chapter 4 discusses resampling techniques in the particle filter approach. State-of-the-art resampling techniques are reviewed, and their limitations are discussed. A novel resampling technique that could keep the PF implementation free of degeneracy while maintaining low level of sample impoverishment is introduced.

Chapter 5 shows the implementation and validation of the developed prognostic model for RUL estimation of IGBT modules under power cycling conditions, which are susceptible to fatigue-induced failures of bond wire lift-off in the package. The results of the novel IMHA-replacement resampling benchmarking other resampling techniques are also presented.

Chapter 6 shows the implementation and validation of the developed prognostic model and resampling techniques for future degradation state estimation of GaN HEMTs under HF-ALT conditions, with the corresponding degradation mechanism of ohmic contact inter-diffusion.

Chapter 7 summarizes the present work and emphasizes again the contributions of this dissertation. Suggestions on the future work are pointed out in this chapter.

Appendix A discusses the investigation on the effect of Gamma irradiation on the DC characteristics of commercial E-mode GaN HEMTs. The DUTs, irradiation procedure and the electrical characterization method are discussed. The DC characteristics of the devices pre- and post-irradiation are compared to gauge the variations of the critical electrical parameters in the irradiations. Potential implementation of the prognostic model developed in the present work is discussed.

Chapter 2: Particle Filter Approach

2.1 Non-linear Bayesian Filtering

The recursive inference and update procedure in dynamic state estimation forms the Bayesian filtering approaches. Generally, dynamic state estimation problems can be expressed in the following equations [99]:

$$x_k = f_k(x_{k-1}, w_{k-1}) \text{ (System Process Model)} \quad (2.1)$$

$$z_k = h_k(x_k, v_k) \text{ (Measurement Model)} \quad (2.2)$$

where x_k is the state variable at time step k , y_k is the observation of x_k at time step k . $f()$ is a nonlinear function of the state variable of the previous state x_{k-1} and system process noise w_k , while $h()$ is a nonlinear function of x_k and measurement noise v_k . w_k and v_k are independent and identically distributed (i.i.d) noise sequences at each time step. They are all random variables following Gaussian distributions.

State estimation based on recursive Bayesian procedure (Bayesian filtering) consists of creating the PDF of the state at each time step k , using the available measurements for the time: $p(x_k|y_{1:k})$, where $y_{1:k} = y_{i:i=1,\dots,k}$. This PDF is created through inference from the previous step and updated with the measurement at the current step. In the Bayesian context, given PDF $p(x_{k-1}|y_{1:k-1})$ at the previous

time step $k - 1$, the prior PDF $p(x_k | y_{1:k-1})$ could be predicted using the Chapman-Kolmogorov equation:

$$p(x_k | y_{1:k-1}) = \int p(x_k | x_{k-1})p(x_{k-1} | y_{1:k-1}) dx_{k-1} \quad (2.3)$$

where $p(x_k | x_{k-1})$ is based on the system process model (Equation 2.1), taking into account the Markov property of x_k . Once a latest measurement y_k becomes available, Bayes theorem can be used to update the prior PDF $p(x_k | y_{1:k-1})$ with y_k , and obtain the posterior PDF $p(x_k | y_{1:k})$:

$$p(x_k | y_{1:k}) = \frac{p(y_k | x_k)p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})} \quad (2.4)$$

where the normalizing constant $p(y_k | y_{1:k-1})$ is obtained from:

$$p(y_k | y_{1:k-1}) = \int p(y_k | x_k)p(x_k | y_{1:k-1}) dx_k \quad (2.5)$$

2.2 Particle Filters

Particle filters are based on Monte Carlo (MC) simulations. Suppose the PDF of the state variable is x_k : $p(x_k | y_{1:k})$, which is assumed to be Gaussian. MC samples could be generated from the distribution, and a discrete approximation of the PDF could be made using the following equation:

$$p(x_k | y_{1:k}) \approx \sum_{i=1}^{N_s} w^i \delta(x_k - x_k^i) \quad (2.6)$$

where i denotes the particle number, w^i is the weight of the i th particle, and x_k^i is the state value of the i th particle, and δ is the Dirac delta function. However, in reality, $p(x_k | y_{1:k})$ may be non-Gaussian which introduces sampling problems. This problem

could be addressed through the method of importance sampling. Suppose that apart from $p(x_k | y_{1:k})$, there exists another PDF $q(x_k | y_{1:k})$ called the importance density, where samples could be easily drawn (e.g. Gaussian PDF). The PDFs $p(x_k | y_{1:k})$ and $q(x_k | y_{1:k})$ would have the following relationship:

$$p(x_k | y_{1:k}) \propto q(x_k | y_{1:k}) \quad (2.7)$$

There is a scaling factor, or a weight between p and q , which varies with x_k :

$$w(x_k) = \frac{p(x_k | y_{1:k})}{q(x_k | y_{1:k})} \quad (2.8)$$

The following equation updates the importance weights stepwise and expand $p(x_k | y_{1:k-1})$ using Equation 2.3:

$$\begin{aligned} \omega(x_k) &= \frac{p(y_k | x_k)p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})q(y_k | y_{1:k})} \\ &= \frac{p(y_k | x_k) \int p(x_k | x_{k-1}, y_{1:k-1})p(x_{k-1} | y_{1:k-1}) dx_{k-1}}{p(y_k | y_{1:k-1}) \int p(x_k | x_{k-1}, y_{1:k-1})p(x_{k-1} | y_{1:k-1}) dx_{k-1}} \end{aligned} \quad (2.9)$$

Assume there is a set of particle state value and weight pairs $\{x_{k-1|k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}$ at time step $k-1$, and the particles are drawn from the importance density $q(x_{k-1} | y_{1:k-1})$, then the posterior PDF at $k-1$ can be approximated by:

$$p(x_{k-1} | y_{1:k-1}) \approx \sum_{i=1}^{N_s} w_{k-1}^i \delta(x_{k-1} - x_{k-1|k-1}^i) \quad (2.10)$$

where w_{k-1}^i is defined as:

$$w(x_{k-1}^i) = \frac{p(x_{k-1}^i | y_{1:k-1})}{q(x_{k-1}^i | y_{1:k-1})} \quad (2.11)$$

At the next time step k , a new set of state value and weight pairs is to be drawn to approximate $p(x_k | y_{1:k})$, the weight update equation could be attained by

substituting Equation 2.10 into Equation 2.9:

$$w(x_k^i) = w(x_{k-1}^i) \frac{p(y_k | x_{k|k-1}^i) p(x_{k|k-1}^i | x_{k-1|k-1}^i)}{q(x_{k|k-1}^i | x_{k-1|k-1}^i)} \quad (2.12)$$

The posterior PDF can then be approximated by:

$$p(x_k | y_{1:k}) = \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_{k|k}^i) \quad (2.13)$$

The recurrence of the above procedure at each time step produces the most basic type of particle filter, which is sequential importance sampling (SIS). The problem of SIS is the particle degeneracy, which will be discussed in the next section.

2.3 SIR Filter

Particle degeneracy refers to that as time step k proceeds, the cumulative effect of process noises results in dispersion of particles x_k , i.e. increasing variance between particles. After a number of steps, all but very few weights tend to be zero, causing the predicted PDF to be predominantly influenced by one or few particles, bringing about significant wastes of computing time and power, as well as loss of generality in the PDF representation [44]. The two general solutions to degeneracy are choosing the proper proposed distribution, and performing resampling [44]. Therefore, some variants of SIS have been developed including the SIR filter [100], the auxiliary sampling importance resampling (ASIR) filter [101] and the regularized particle filter (RPF) [102], developed to solve the issue of degeneracy in SIS filter [44]. A classic example of using resampling to solve the degeneracy issue is the SIR filter, which is the most prevalent filter used in prognostics [103]. The steps of the SIR filter are as follows [104]:

- *Initialization:*

Let k be the number of time steps in sequence. To initialize, set $k = 0$ and generate a set of N particles $x_0^i, i = 1, 2, \dots, N$ from the prior distribution of the state variable x at $t = 0$, denoted $p(x_0)$.

- *Importance Sampling and Calculation of Weights:*

At each time step k , generate a set of particles x_k^i from the proposal distribution $q(x_k^i | x_{0:k-1}^i, z_{0:k})$, where $z_{0:k}$ are the series of measurements of the state. In the SIR filter, define $q(x_k^i | x_{0:k-1}^i, z_{0:k}) = p(x_k^i | x_{k-1}^i)$, which is the conditional distribution of x_k^i given the set of particles in the previous step x_{k-1}^i . Assign each particle a weight based on Equation 2.12, and normalize the weights.

- *Resampling:*

If the effective sample size N_{eff} computed using Equation 2.14 is below the given threshold N_{th} , then resampling is performed. Commonly, $N_{th} = 2/3N$ is taken as the threshold.

$$N_{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2} \quad (2.14)$$

Generate N new particles \tilde{x}_k^i from the current particle set x_k^i , and replace x_k^i with the new set. The weights of the new set are assumed to be equally distributed:

$$\tilde{\omega}_k^i = \frac{1}{N} \quad (2.15)$$

- *State Prediction:*

Calculate the estimated state \hat{x}_k using Equation 2.13.

2.4 Summary

This chapter has given an overview of dynamic state estimation problems and expounds on the applicability of non-linear Bayesian filters in solving these problems. The theoretical background of particle filters in general, and the SIR filter has been presented. The present dissertation uses the SIR filter method to perform RUL prediction of power electronic devices, which will be presented in the next chapter.

Chapter 3: Prognostic Model Development

3.1 Failure Precursors and Failure Thresholds

Bayesian tracking problems require a variable whose time-series evolution can indicate the real-time health status of the system under a specific type of stressor, which eventually results in specific failure modes. This variable, called failure precursor in the context of reliability engineering, is the basis of condition monitoring and RUL prediction, and therefore requires careful selection as the first step of prognostic model formulations [78]. Ideal failure precursors should be directly gaugeable using metrology methods in order to minimize the modeling and measurement errors introduced in the conversion process. Failure threshold is another important parameter to be defined. In many cases, even if the systems have not lost its function completely, the degradations of performance may have been beyond an acceptable minimum level. This level can be identified as the failure threshold, beyond which the system would be considered failed even though still functional. In other words, parametric failure is introduced as the failure criterion in place of functional or catastrophic failure.

As an IGBT is turned on, current flows from the collector to the emitter, denoted I_C . In an IGBT module, the current flows not only through multiple

semiconductor die but also through passive components as well as interconnects between components connected via on-chip bond wires. The net result is an on-state forward voltage drop $V_{forward}$ between the collector-side and emitter-side terminals of the module. In power cycling tests, the current from collector to emitter is kept constant, and hence as cracks develop near the interface of bond wires and bond pads, the on-state resistance R_{on} of the module increases, resulting in an increase in $V_{forward}$. Therefore, $V_{forward}$ has commonly been used to as a failure precursor of IGBT modules and it was selected as the failure precursor in the present work as well. The threshold of device failure was a 5% increase in $V_{forward}$ as was defined in Lutz et al. [78]. This is an empirically defined threshold that is considered an appropriate point for immediate actions to be taken on devices. Large increases in $V_{forward}$ can be anticipated after this threshold since the number of bond wires detaching increases. The functional failure of devices under test (DUTs) occurs when enough bond wires detach to completely cut off the electrical connection in the module.

For many target applications, GaN HEMTs under high-frequency (HF) drive conditions operate at a nominal drain voltage (most typically 28 V, 40 V and 50 V) and a high gain compression level (decrease in gain due to nonlinearity of transfer function in amplifying devices [105]) due to high input power [106]. It was discussed in Section 1.6 that HF GaN HEMTs were susceptible to drain current reduction and power output degradation due to defects in the active device layers trapping electrons under high vertical electric field near the gate. Therefore, power output P_{out} was taken as the failure precursor for the degradation of GaN HEMTs and was

due to ohmic contact inter-diffusion investigated and identified in this dissertation. As for the failure threshold, a parametric criterion of -1 dB in power output was defined in the original paper [69]. However, by the end of test, the degradation still did not reach this threshold. i.e. The pre-defined failure did not occur. Therefore, the actual P_{out} level at the end of test was used as a reference for the future state prediction accuracy of GaN HEMTs in the present dissertation.

3.2 Model Framework

It was mentioned in Section 2.1 that dynamic state estimation problems involved a system process model and a measurement model. They were essential elements in the particle-filter-based RUL prediction model. The prognostic model proposed in the present work integrated an anomaly detection module and a RUL prediction module. The RUL prediction module consisted of a non-linear process model and a measurement model. The implementation of prognostic model included three stages: a nominal stage, a training stage and a RUL prediction stage. The nominal stage was an initial stage that establish a "nominal region" in the raw failure precursor data series, including all the healthy data to establish the baseline for anomaly detection via feature construction, feature extraction and clustering on the data instances in the moving window. After the baseline was established, the training stage worked on the subsequent failure precursor data instances beyond the upper bound of the "nominal region". As a new measurement of the failure precursor became available at each time step, the moving window was shifted to include this new measurement

and drop the oldest one. Then feature construction and feature extraction were implemented on the data instances in the moving window. The affiliations of the obtained data instance in the feature space to the established baseline clusters were determined, by assessing the outlier probability of this data instance for each cluster. Simultaneously, the recursive Bayesian process was applied to the process model, updating the model parameters as well as the particle weights and weights with sequential new de-noised measurements. The measurement model performed pre-processing of the raw measurement, supplying the de-noised measurement to the process model. Overall, the anomaly detection module and the Bayesian filter in the RUL prediction module examined test data instances for anomaly and updated the process model in parallel throughout the training stage. Once an anomaly was detected, the training stage terminated and the RUL prediction stage began. Updates of the process model and particles ceased, and RUL prediction was triggered, performing unremitting stepwise future state estimation based on the process model and the particle positions/weights at the anomaly signaling point (ASP).

3.3 Anomaly Detection Technique

3.3.1 Feature Space

In machine learning, a feature is an individual measurable property or characteristic of an observed phenomenon [107]. In data analytics, it is a attribute that describes the inherent characteristics of raw data instances [108]. The attribute can be of various types including binary, categorical, continuous, and so on [41].

Selecting a set of informative, discriminating and independent features is crucial for the effectiveness of many machine learning techniques including regression, classification and clustering. A feature vector is a multi-dimensional vector of typically numerical features representing an object, and the corresponding multi-dimensional space where objects represented by feature vectors reside in is the feature space. A high-dimensional feature vector feature space could consist of redundant and correlated features. Therefore, dimensionality reduction needs to be implemented to obtain a smaller set of features that are independent and enlarge data variance, constituting an optimized feature space.

3.3.2 Nominal Region

The benefits of semi-supervised machine learning have been discussed in Section 1.3, and these benefits were adopted in the present work for anomaly detection. Semi-supervised anomaly detection requires establishing a set of healthy or nominal data representing the expected behavioral patterns of the failure precursor, as the baseline for anomaly detection. In the present work, the available raw data were time-series degradation data of power electronic devices without any labelled healthy data. Nevertheless, a complete data series recorded the entire degradation process of the devices from initial healthy status to eventual failure. It is possible to select an initial part of the data series and define it as a “nominal region”. All the data instances within this region were considered healthy data to establish the baseline for anomaly detection.

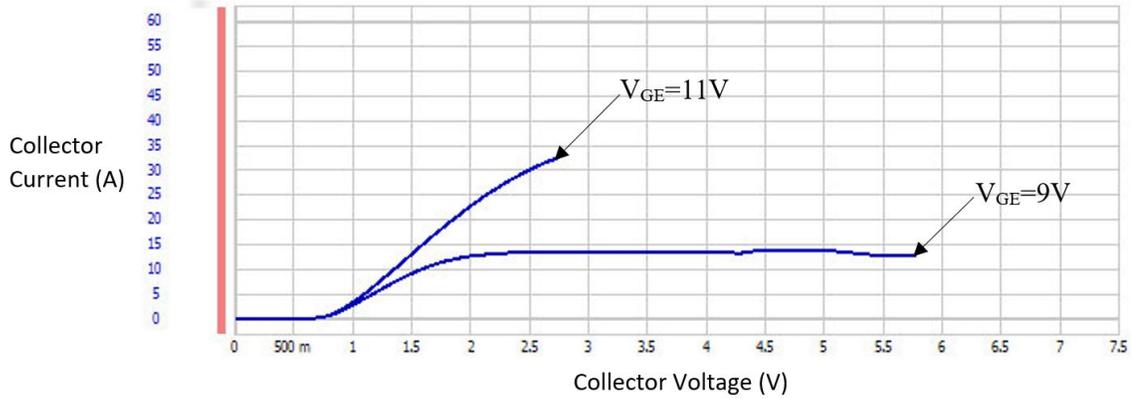


Figure 3.1: Example of I-V output characteristics curves of an IGBT in Infineon FS20R06W1E3, given a gate voltage V_{GE} of 9 V and 11 V, respectively

The “nominal region” in the raw data was determined experimentally through evaluating the variations of the I-V output curves of 60 IGBTs in 10 IGBT modules, characterized using an Agilent B1505A curve tracer. The characterized IGBT modules were Infineon FS20R06W1E3 with a rated collector-emitter voltage (V_{CE}) of 600 V [5], and each module consisted of 6 IGBTs. The measured output curves with a gate voltage V_{GE} of 9 V and 11 V for one of the IGBTs, is shown in Figure 3.1. The point on these I-V curves selected for distribution analysis was $I_C = 30$ A, $V_{GE} = 11$ V. By collecting the corresponding V_{CE} values of all the 60 IGBTs, a sample set with a size of 60 was established and imported into MINITAB for parametric normal distribution analysis. The normal distribution was selected to fit this data set since it outperformed the Weibull distribution in goodness-of-fit for the dataset. The 95% confidence interval (CI) of the mean of the fitted normal distribution was considered an acceptable variation interval for healthy data. This was because the mean of the distribution, calculated based on a relatively large population, could be

taken as the estimated $V_{forward}$ output of a single “nominal device”. The 95% CI of the mean could thus be considered the acceptable variation interval of this “nominal device”. MINITAB calculations showed that the 95% CI of the mean of V_{CE} at the specified point was [2.496 2.529]. Therefore, it was assumed that the data instances whose values stayed within $\pm 0.6546\%$ of the nominal value were considered within the “nominal region”. This range was applied to both of the power devices studied in the present work.

3.3.3 Feature Construction and Extraction

The semi-supervised anomaly detection module in the present dissertation utilized unsupervised machine learning techniques to establish baseline clusters for the healthy data within the “nominal region”, including principal component analysis (PCA) for dimensionality reduction and k -means clustering. The first step was feature construction by selecting and calculating a number of features that possibly contain useful information to discriminate between normal and anomalous data instances. Nine common time-domain statistical features, given below, were selected and calculated at each time step t by setting a moving window that included the 10 latest $V_{forward}$ data instances. Note that s denotes the failure precursor, or $V_{forward}$.

- Mean $\overline{ds/dt}$
- Maximum $(ds/dt)_{max}$
- Root mean square $(ds/dt)_{rms}$

- Standard deviation $\sigma_{ds/dt}$
- Crest factor $CF_{ds/dt}$
- Impulse factor $IF_{ds/dt}$
- Shape factor $SF_{ds/dt}$
- Energy of the domain $ED_{ds/dt}$
- Clearance factor $CLF_{ds/dt}$

Dimensionality reduction was then performed using PCA. Dimensionality reduction is an important task in unsupervised machine learning involving either feature extraction to compute some new features or feature selection to select the best original features. Feature extraction was adopted in the present work to obtain a small set of new features that captured most of the variance in the raw data, which were called principal components (PCs). In this way, the redundancies or correlations in the preliminarily constructed statistical features were removed. In fact, the essence of PCA is maximizing the variances of the features along the axis of the PCs in the feature space to retain and intensify useful information, while removing the correlations between pertinent features to dislodge the redundant information.

Mathematically, PCA is described as an orthogonal linear transformation that converts the data to a new coordinate system (PC-based feature space) such that the greatest variance by any projection of the data could lie on the first coordinate, called the first PC, while the second greatest variance on the second coordinate, or the second PC, and so on [109]. The original data matrix X , which is a collection

of vectors representing random variables, is multiplied by a matrix V consisting of orthonormal basis vectors, or eigenvectors, to get the extracted data matrix called the score matrix Y :

$$Y = XV \tag{3.1}$$

The orthonormal basis matrix V can be obtained using singular value decomposition (SVD). SVD is a matrix factorization step decomposing the original data matrix into three new matrices:

$$X = U\Lambda V^T \tag{3.2}$$

where U is a unitary matrix, Λ is a diagonal matrix with non-negative real numbers on the diagonal line, which are the singular values. In the present work, PCA was implemented on the healthy data instances in the “nominal region”, which served as the training data. The number of PCs was determined using the criterion that the cumulative percentage of variance that all the PCs captured just exceeded 80%. For the time-series degradation data studied in the present work, the number of PCs varied from 2 to 4. The PC-based feature space could hereby be established, and the converted raw data instances could be located in the feature space.

3.3.4 Singular-Value-Weighted Distance

In anomaly detection, particularly classification and clustering, an appropriate proximity measure between data instances, typically a distance or similarity matrix, is required to differentiate between healthy and anomalous data instances in the feature space. The choice of the proximity measure is critical to the performance

of the technique. The most commonly used distance measures are the Euclidean distance (ED) and the Mahalanobis distance (MD). The advantage of MD over ED is that it is based on the covariance among variables in the feature vectors, utilizing the mean and variance of variables. Therefore, the scale and correlation issues in the Euclidean distance are eliminated [110]. Nevertheless, in the present work, PCA already conducted normalization and correlation elimination on the raw data instances, so there was no clear advantage of using MD in this case. ED is easier to compute and interpret than MD [111], so it was adopted and further improved, creating a singular-value-weighted distance (SVWD) measure used in the present work. The conventional ED, $d_i(k)$, in the PC-based feature space is given in Equation 3.3:

$$d_i(k) = \sqrt{(PC_1(k) - PC_{1m}^j)^2 + (PC_2(k) - PC_{2m}^j)^2 + \dots + (PC_n(k) - PC_{nm}^j)^2} \quad (3.3)$$

where PC_1, PC_2, \dots, PC_n denote the coordinates of the test data instance in the n -dimensional feature space, given that the number of PCs is n ; $PC_{1m}^j, PC_{2m}^j, \dots, PC_{nm}^j$ denote the coordinates of the mean of the baseline cluster in the feature space. The variance each PC accounts for is indicated by their corresponding singular values. Conventional ED considers all the PCs as equally weighted. The idea of creating SVWD is that even though the variance has been accounted for in the PC variables themselves, further attenuating the PCs capturing more pronounced variances may help differentiate normal and anomalous data even more clearly. Therefore, the conventional Euclidean distance was revised by assigning the corresponding singular values as weights to the PCs, to further intensify the “more important” PCs, as is

shown in Equation 3.4:

$$d_i(k) = \sqrt{\left(\frac{SV_1(PC_1(k)-PC_{1m}^j)}{SV_1+SV_2+\dots+SV_n}\right)^2 + \left(\frac{SV_2(PC_2(k)-PC_{2m}^j)}{SV_1+SV_2+\dots+SV_n}\right)^2 + \dots + \left(\frac{SV_n(PC_n(k)-PC_{nm}^j)}{SV_1+SV_2+\dots+SV_n}\right)^2} \quad (3.4)$$

where $SV_i (i = 1, 2, \dots, n)$ are the singular values corresponding to each PC.

3.3.5 K -means Clustering

Clustering is the process of partitioning a set of data (or objects) into a set of groups called clusters based on some similar characteristics [112]. It is an important technique of exploratory data analysis, aimed at investigating the internal structure of complex data sets that could not be sufficiently exploited by basic statistics such as the mean and the covariance [113]. Typical applications of clustering include pattern recognition, image processing, spatial data analysis, fault detection, etc. Partitioning-based clustering is one type of clustering dividing the data set into a desired number of clusters randomly, and iteratively relocating data points between clusters based on a specific criterion until an optimal partition or a stopping criterion is met [113]. K -means clustering is the most typical partitioning-based clustering technique, partitioning the original data set into k clusters based on the features and the specified proximity measure.

The clustering process is implemented by the following three steps:

1. Randomly select k data points as centroids, where k is the desired number of clusters.
2. For the rest of data instances, each of them is assigned to the nearest centroid, based on the selected distance measure (SVWD in the present work).

3. As the assignments are complete for all instances, the k centroids are recomputed. This is done by taking the mean of all data instances assigned to the cluster.

The clustering process kept iterating steps 2 and 3 until a stopping criterion is met. This criterion could be no data points changing clusters anymore, or the sum of the distances being minimized, or the maximum allowed number of iterations being reached.

In the present work, an optimal number of baseline clusters were established for the healthy data instances in the “nominal region” using k -means clustering, based on the SVWD measure developed. The number of clusters, or the value of k , was selected using the average silhouette method, which was introduced by Kaufman and Rousseeuw in 1990 as an effective and direct method for determining an optimal number of clusters based on comparison of the tightness and separation of the clusters in each case [114].

3.3.6 Anomaly Probability Calculation

With the baseline clusters established, test data instances, which were all the data instances not included in the “nominal region”, were examined by computing their outlier probability for each of the baseline clusters. The same moving window containing the 10 latest data instances was also implemented to compute the statistical features. Then the test data instances were multiplied by the eigenvectors, locating

them in the PC-based feature space:

$$Y_i = X_{test} V_i \quad (3.5)$$

where Y_i denotes the i th PC, X_{test} denotes the vector of test data instance (computed statistical features), and V_i denotes the i th eigenvector.

Conventional K-means clustering approach determines the affiliation of a test data instance to a baseline cluster by computing the distance between the test data point and the centroid of the cluster [115]. The advantages of this method are simplicity and fast computation speed. However, a significant issue with the conventional approach is its sensitivity to outliers. A conspicuous outlier in a cluster could cause remarkable shift in the position of the centroid, therefore providing a distance between the test data point and the centroid deflected compared with the majority of points in the cluster. A new improved approach more robust to outliers is necessary, offering a statistics-based estimate of the distance between the test data instance and the “nominal region” to detect anomalies. Besides using SVWD as distance measure, another novelty of the anomaly detection module in the present work lies in using Fisher criterion (FC), which is the gauge of the relative separation between the distributions of two classes commonly used in Fisher’s linear discriminant analysis. An anomaly was considered detected when a test data instance failed to conform to any baseline cluster. For each baseline cluster, the set of distances between different healthy data instances within the cluster can be computed. Then for each test data instance examined, the set of distances between the test data instance and the healthy data instances in the baseline cluster can also be computed.

Consequently, two distributions of distances were obtained corresponding to each baseline cluster and each test data instance. The idea of determining the outlier probability arose from justifying whether these two distributions, representing two sets of observations, could be considered belonging to the same distribution to a certain degree of confidence, to assess whether the examined test data instance belongs to the corresponding cluster. This is exactly how FC functions. In reliability engineering, FC is identified as the safety margin (SM), indicating the degree of relative separation between the means of two distributions (commonly load and strength) [116]. Unless it belonged to any baseline cluster, the test data point was considered an outlier. Therefore, FC values for each data instance corresponding to all the baseline clusters were calculated, assuming Gaussian distributions. The equation of FC is as follows:

$$FC_x = \frac{\mu_{test} - \mu_{healthy,x}}{\sqrt{\sigma_{test}^2 + \sigma_{healthy,x}^2}} \quad (3.6)$$

where μ_{test} , σ_{test} are the mean and the standard deviation of the distribution of d_t between the test data instance and healthy baseline cluster x ($x = 1, 2, \dots, k$, where k is the number of baseline clusters), respectively, and $\mu_{healthy}$, $\sigma_{healthy}$ are the mean and the standard deviation of the distribution of distances within cluster x , respectively. The attained FC value could be converted to the probability of the test data instance not belonging to the baseline cluster x , i.e. the outlier probability, denoted by R_x :

$$R_x = \Phi(SM_x) \quad (3.7)$$

A critical reliability threshold of $R_x = 0.995$ was defined. Once the R_x values for

all the baseline clusters exceeded 0.995, i.e. the test data instance did not belong to any healthy cluster with at least 99.5% confidence level, the examined test data instance would be announced an anomaly.

3.3.7 Validation of Anomaly Detection Technique

This section presents the validation of the proposed anomaly detection technique discussed in the preceding sections. It is difficult to evaluate the outcome of unsupervised/semi-supervised machine learning without a benchmark example whose actual outcome is explicitly given [117]. Therefore, to evaluate the robustness of the proposed technique, a benchmark dataset with binary outputs of 0 or “inlier” and 1 or “outlier” is required. One renowned dataset for evaluating clustering algorithms is the Wine dataset from UCI machine learning repository [118]. The dataset is a multiclass classification dataset with 13 attributes and 3 classes, containing the chemical analysis results of wines from three different cultivars in a region of Italy [119]. The quantities of 13 constituents in the three types of wines were recorded. The instances from two of the classes were used as inliers and the ones from the third class containing 10 instances were taken as outliers.

The validation of the proposed technique also required performance comparison with state-of-the-art techniques, such as the Mahalanobis Distance (MD) approach. The MD approach has been extensively used in prognostics. For example, in electronics, it has been implemented for detecting anomalies in IGBTs [50] and multilayer ceramic capacitors [120]. The MD-based technique as was introduced

in [50] followed a similar procedure as the proposed technique, but with a different criterion of anomaly determination:

- Perform k-means clustering based on MD for each baseline cluster and obtain a distribution of MDs between healthy data instances within the cluster. Compute the mean and standard deviation of the distribution.
- Compute the MD between the current test data instance and the centroid of each baseline cluster.
- Compute the Z -score of the MD value for the test data instance corresponding to each baseline cluster. If $Z - score > 3$ or $Z - score < -3$ (other thresholds may apply), then the test data instance is considered not belonging to the cluster. If this criterion holds for all the healthy clusters, then the test data instance is considered an anomaly.

For each anomaly detection technique, the implementation results on the Wine dataset were compared with the true classification results using Rand Index (RI), proposed by William M. Rand in 1971 [121], has been widely used in statistics as a measure of similarity between two data clusters. Assuming there are a set of n objects S for partitioning, and two partitioning results, $X = X_1, \dots, X_i$, and $Y = X_1, \dots, X_j$, are available which partition the n objects into i and j subsets, respectively. Suppose that:

- a is the number of pairs of elements in S that are in the same subset in both X and Y .

- b is the number of pairs of elements in S that are in different subsets in both X and Y .
- c is the number of pairs of elements in S that are in the same subset in X but in different subsets in Y .
- d is the number of pairs of elements in S that are in different subsets in X but in the same subset in Y .

Then obviously $a + b$ gives the total number of element pairs that yields the same clustering results in both methods. The Rand index R is given as the proportion of these pairs in all the pairs, as is shown below:

$$R = \frac{a + b}{a + b + c + d} \quad (3.8)$$

The denominator is the total number of element pairs, so for n total elements:

$$a + b + c + d = \binom{n}{2} = \frac{n(n-1)}{2} \quad (3.9)$$

Therefore, the Rand index R is a measure of agreement between the two clustering results that lies between 0 and 1, with 0 being total disagreement, and 1 being total agreement.

The problem of the conventional Rand index is that the expected value of two random clustering results is not a constant one [122]. To address the problem, the adjusted Rand index (ARI) was proposed by Hubert and Arabie in 1985 [123], by using the generalized hypergeometric distribution as the random model. Specifically, the clustering results X and Y are randomly selected so that the number of objects

in the clustering results is fixed. A baseline is established by using the expected level of agreement of all element pairs between clustering results [124]. Therefore, ARI has become a more robust measure of similarity than the conventional RI. Equation 3.10 shows the equation for computing ARI based on the expected index (level of agreement). In the present work, the ARI values of the proposed anomaly detection technique and the MD-based technique shown above were compared across a variety of thresholds for the outlier probability for each baseline cluster.

$$ARI = \frac{Index-Expected\ Index}{Max\ Index-Expected\ Index} = \frac{\sum_{ij} \binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right] / \binom{n}{2}} \quad (3.10)$$

Table 3.1 summarizes the results of ARI values of two techniques across thresholds varying from 90% to 99.73% (upper bound 6σ). It turned out that the proposed technique in the present work demonstrated higher ARI values than the MD-based technique for all the thresholds from 90% to 99.5%, with the maximum value incurred at 99.5%. This also explained why 99.5% was selected as the critical reliability threshold in the previous subsection.

Table 3.1: Results of adjusted Rand index (ARI) Comparison for the proposed anomaly detection technique

Probability Threshold	Adjusted Rand Index	
	Proposed Technique	MD-based Technique
90%	0.3234	0.2586
95%	0.3402	0.2586
99%	0.3303	0.303
99.5%	0.35	0.303
99.73%	0.2096	0.303

3.4 Bayesian Filter Model

3.4.1 Physics-of-Failure-Based Model Variables for IGBT Modules

As was presented earlier, the RUL prediction model consisted of a system process model and a measurement model. The process model is the basis of future state and RUL prediction, so its development is critical to the performance of the prediction model. Generally, a non-linear mathematical model is required to describe the evolution behavior of the failure precursor, because of the non-linearity of failure precursor evolutions. The failure precursor of IGBT modules, $V_{forward}$, evolves slowly at first and then gradually accelerates due to increasing pace of fatigue-induced crack propagation in the bond wires. The time-series evolution of at least one variable composing the process model should be able to effectively capture this nonlinearity

of degradation, to correctly “drive” the evolution of the failure precursor. The most straightforward choice of this “driving variable” is t , the elapsed test time or on-power cycles, as was used in [125], [126]. Obviously, the problem of using t is lack of non-linearity, so additional variables are required to make up for it.

An alternative approach is to use the physics-of-failure model associated with the relevant failure mechanism, given that a strong correlation exists between the failure precursor and the magnitude of degradation or damage in the system due to the mechanism. The selected “driving variable” should be correlated to the damage magnitude while calculable at each time step. Hereby, the failure precursor parametric degradation could be related to the physical degradation evolution of the devices. For IGBT modules under the power cycle loading condition with fatigue-induced bond wire lift-off as the dominant failure mechanism, the direct damage scale is the crack size near the bond wire-bond pad interface. The first principle of this failure mechanism is the well-known Paris equation. In fact, the complete failure physics behind bond wire lift-off is fatigue-induced void and crack initiation, which is a much more complex procedure than that of Paris-Equation-based crack propagation. However, a complete three step fatigue model is much more difficult to set up for the fault detection process since the crack initiation is based on knowledge of pre-existing material defects in the bond wire or bond pad. Therefore, only crack propagation part was considered in the present work.

Besides the “driving variable” addressing the nonlinearity of failure precursor evolution, variables accounting for other factors influencing the speed of degradation, such as the effect of loads or stresses on the tested devices, should also be included

in the process model. Orchard et al. [29] used a hypothetical loading parameter incorporating in-situ loading conditions, which is a useful point because load or stress level is a critical factor influencing the speed of degradation. In the present dissertation, a hypothetical loading variable is also created, but it is more clearly defined by being analytically associated with critical loading parameters of the stress test.

Specifically, for IGBT modules, two time-dependent variables $l(k)$ and $\Delta K(k)$ were developed and accommodated in the process model, in which k denotes the present time step of cumulative power cycles. $l(k)$ is a hypothetical variable associated with the power cycle loading parameters, the load current I_{load} , and the junction temperature variation ΔT_j , as is shown in Equation 3.11. The values of ΔT_j were derived from the directly measurable base plate temperature, which was monitored in-situ throughout the power cycling tests in parallel with $V_{forward}$. In this sense, $l(k)$ tracks the real-time loading condition in the test.

$$l(k) = f_l(I_{load}, \Delta T_j(k)) \quad (3.11)$$

$\Delta K(k)$ is the stress intensity factor governing crack propagation under fatigue loads. The reason of choosing this variable is that in the theory of fatigue, $\Delta K(k)$ is correlated to the crack length a and the stress amplitude $\Delta\sigma$, as Equation 3.12 indicates. Through some derivations, it could become an explicit function of the elapsed power cycle N .

$$\Delta K(k) = \alpha \Delta\sigma \sqrt{\pi a} \quad (3.12)$$

where α is a constant. $\Delta\sigma$ is related to ΔT_j , while the crack length is not directly

calculable. The Paris Equation for fatigue crack extension is:

$$\frac{da}{dN} = A (\Delta K)^n \quad (3.13)$$

where A , n are material-related constants. Combining Equation 3.12 with 3.13 and performing an integral over ∂a and ∂N , respectively, yielded the relationship between $\Delta K(k)$ and N , the elapsed power cycles, given in Equation 3.14. The bond wire material was specified to be aluminum. The NASGRO database [127] gave the n value of around 3.2 for aluminum and aluminum alloys.

$$\Delta K(k) \propto B \left[A \left(1 - \frac{n}{2} \right) B^n N + a_i^{1-\frac{n}{2}} \right]^{\frac{1}{2-n}} \quad (3.14)$$

where $B = 1.12\pi\Delta\sigma$, in which $\Delta\sigma$ is the stress amplitude at the crack location, and in turn associated with the junction temperature variation ΔT_j with the following relationship:

$$\Delta\sigma = CE (\alpha_{Al} - \alpha_{Si}) \Delta T_j \quad (3.15)$$

where $C < 1$ is a constant, E is the young's modulus of aluminum, and α_{Al} and α_{Si} are the CTEs of the aluminum and silicon, respectively. An example of the resultant time-series data of $\Delta K(k)$ for degradation data series in [8] is shown in Figure 3.2.

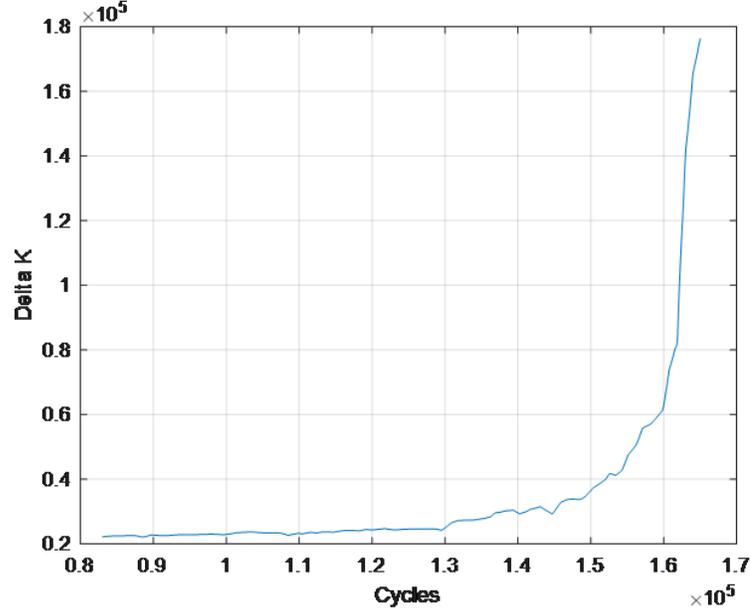


Figure 3.2: Time-series data of $\Delta K(k)$ for degradation data series in [8]

3.4.2 Physics-of-Failure-Based Model Variables for GaN HEMTs

For HF GaN HEMTs, the device structure, operating condition and failure mechanisms are completely different from those of IGBT modules. The output power P_{out} , as the failure precursor of GaN HEMTs under RF-ALT, decreases more rapidly at the beginning stage of the test, as the ohmic contact diffusion rate is the fastest at the beginning stage, and the defect generation rate in the active layers is also the fastest, trapping a large number of electrons. As tests go on, the concentration of Au atoms in the source contact increases, and as a result the diffusion rate becomes slower. All in all, the evolution of P_{out} is also nonlinear. Similar to IGBT modules, the system process model of GaN HEMTs also consists of a hypothetical loading parameter, and a physics-of-failure related variable.

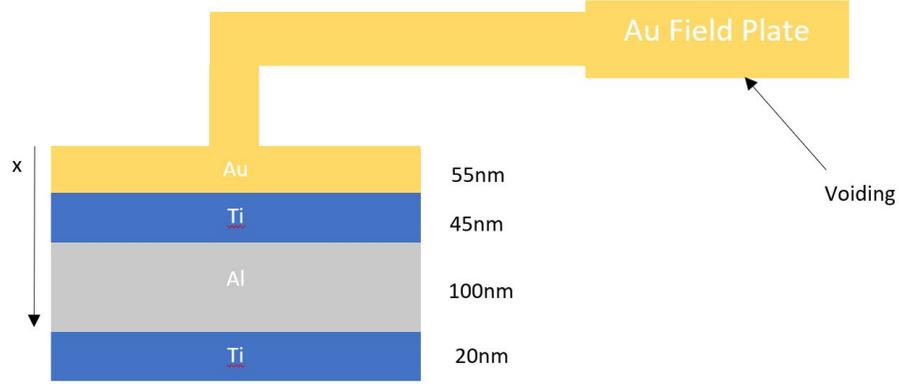


Figure 3.3: Source contact stacked layers and source-connected field plate of Wolf-speed GaN HEMTs

The “driving variable” for GaN HEMTs is associated with ohmic contact diffusion that leads to voiding in the Au field plate, which in turn leads to loss of function of the field plate. Figure 3.3 illustrates the structure of the source contact and the field plate physically connected to the source. It was assumed that the ohmic contact had a stacking layer structure of Ti/Al/Ni/Au (200/1000/450/550 Å) [128]. The length of the source contact was $3\ \mu\text{m}$, while the width of the device was $3.6\ \text{mm}$ [129]. As the inter-diffusion occurred, Au atoms diffused into the source contact consisting of Ti/Al/Ti stacking layers, while the Ti and Al atoms were believed to diffuse to the top of the Au layer in the source contact. Therefore, this problem was becoming an interdiffusion problem between Au, Ti and Al. The void size, denoted $lv(t)$, in the field plate was the right candidate for the ”driving variable” established on a well understood and formulated first-principle model, Fick’s Second Law of diffusion:

$$\frac{\partial C(x,t)}{\partial t} = \frac{\partial}{\partial x} \left(\frac{\tilde{D}(C) \partial C(x,t)}{\partial x} \right) \quad (3.16)$$

where $C(x, t)$ denotes the concentration of the diffusing species (in this case Au) at position x and time t . $\tilde{D}(C)$ denotes the inter-diffusivity between diffusing species and host material, which is a function of elapsed test time as well. The void size can be microscopically interpreted as the total number of Au atoms from the field plate diffusing into the Ti/Al/Ti layers, which is computable from the concentration profile. The rate of diffusion and void size are related to the geometry of the GaN HEMT, particularly the dimensions of source contact and field plate.

To solve Equation 3.16, $\tilde{D}(C)$ had to be determined. Diffusivity is dependent on temperature. Therefore, a thermal simulation of the investigated Wolfspeed GaN HEMT was necessary to identify the local temperature at the source contact. A 2D CAD model of the device was created using COMSOL Multiphysics, simplified over the device cross-section plot provided in [130], taking the device width perpendicular to the cross section into account. The material of each layer provided in [129], [130] was defined and assigned. The electrical loads $V_d = 50\text{ V}$ as specified in [69] were applied to the device. Then Electrical Current (EC), Electrostatics (ES) and Heat Transfer in Solids (HT) analysis modules were invoked to simulate the thermal profile of the device during the RF-ALT. Figure 3.4 shows the temperature profile of the device under steady state during the test. The local temperature can be extracted from the profile for any point of interest. The diffusion temperature extraction point, which was located inside the source contact, is marked in the figure as well. It turned out that the temperature at this point was around 302°C , and had little variation along the length of the source contact. Therefore, this temperature was used in the diffusivity computation to be discussed subsequently.

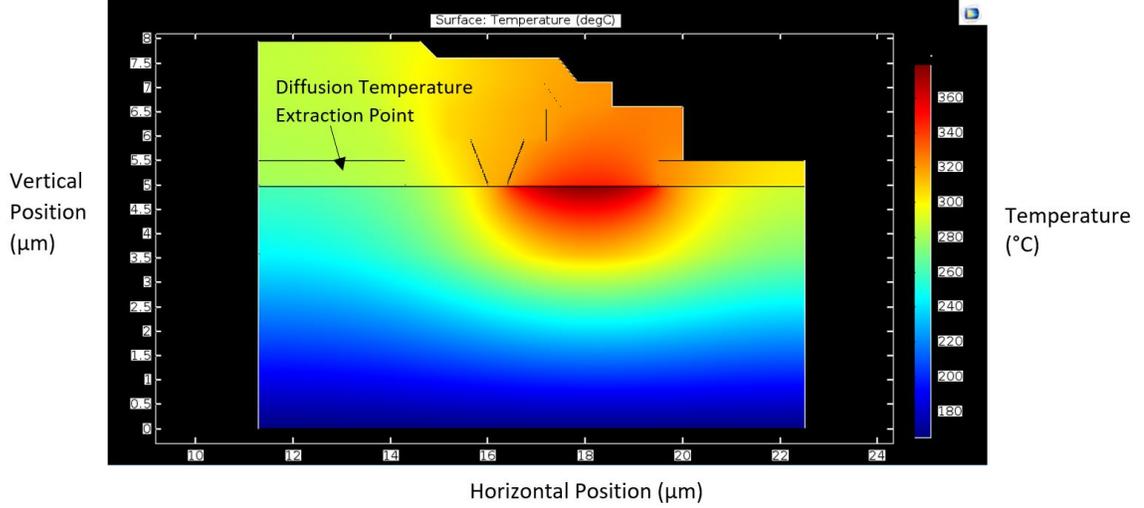


Figure 3.4: Temperature profile simulation of Wolfspeed GaN HEMT under RF-ALT

In addition, the inter-diffusivity in each layer in the source contact differed with the material of each layer. To derive a solution to the diffusion equation, the inter-diffusivities between Au, Ti and Al needs to be determined. The Darken-Manning Equation, which links the interdiffusion coefficient, the tracer diffusion coefficients and the thermodynamic factor, was used [131]. Equation 3.17-3.19 shows the Darken-Manning Equation the three lower source contact layers (upper Ti layer, Al layer and lower Ti layer):

$$\widetilde{D}_1 = \{c_{Ti@Ti_1}D_{Au} + c_{Au@Ti_1}D_{Ti}\} S\varphi \quad (3.17)$$

$$\widetilde{D}_2 = \{c_{Al@Al}D_{Au} + c_{Au@Al}D_{Al}\} S\varphi \quad (3.18)$$

$$\widetilde{D}_3 = \{c_{Ti@Ti_2}D_{Au} + c_{Au@Ti_2}D_{Ti}\} S\varphi \quad (3.19)$$

where \widetilde{D}_n ($n = 1, 2, 3$) denote the interdiffusion coefficients in the three layers, $c_{Ti@Ti_1}$, $c_{Au@Ti_1}$, \dots , etc. are the size fractions of each component in the corresponding layer, in which Ti_1 , Ti_2 denote the upper and lower Ti layer, D_{Au} , D_{Ti} and D_{Al} denote

the tracer diffusion coefficient of Au, Ti and Al, respectively, S denotes the Manning factor and φ denotes the thermodynamic factor. The Manning factor can be ignored generally because the Manning kinetics theory puts a fairly narrow range on S , which is very close to 1 [132]. φ is defined as

$$\varphi = \frac{c_{Au}}{kT} \frac{\partial \mu_{Au}}{\partial c_{Au}} \quad (3.20)$$

where μ_{Au} is the chemical potential of Au. φ can be expressed in terms of the activity as

$$\varphi = 1 + \frac{d \log f_{Au}}{d \log c_{Au}} \quad (3.21)$$

where f_{Au} is the activity coefficient of Au. The activity vanishes for ideal fluid mixtures, and the thermodynamic factor is equal to unity in this case. Therefore, in the present work, it is assumed that the enthalpy of mixing is negligible, so the mixture of Au and Al can be considered almost ideal, and $\varphi \approx 1$ [133]. Hence, the inter-diffusivities in Equation 3.22-3.24 can be approximated as

$$\widetilde{D}_1 = c_{Ti@Ti1} D_{Au} + c_{Au@Ti1} D_{Ti} \quad (3.22)$$

$$\widetilde{D}_2 = c_{Al@Al} D_{Au} + c_{Au@Al} D_{Al} \quad (3.23)$$

$$\widetilde{D}_3 = c_{Ti@Ti2} D_{Au} + c_{Au@Ti2} D_{Ti} \quad (3.24)$$

At the beginning of diffusion ($t = 0$), for Al and Ti layers, $c_{Au} = 0$, $c_{Al} = c_{Ti} = 1$. Thus $\widetilde{D}_1 = \widetilde{D}_2 = \widetilde{D}_3 = D_{Au}$. The tracer diffusion is usually assumed to be identical to self-diffusion (assuming no significant isotopic effect) [134]. The temperature dependence of the self-diffusion coefficients is

$$D = D_0 \exp\left(-\frac{Q_d}{RT}\right) \quad (3.25)$$

where D_0 is a pre-exponential coefficient, Q_d is the activation energy for diffusion, R is the gas constant, and T is the temperature. The D_0 and Q_d of Au, Ti and Al can be found in [135]. Based on Equation 3.25, the corresponding coefficient, and the temperature $T = 302^\circ\text{C}$, D_{Au} , D_{Ti} and D_{Al} can be obtained. Table 3.2 summarizes their values:

Table 3.2: Summary of self-diffusion coefficients of interdiffusion components

Material	Diffusion Coefficient (m^2/s)
Au	1.44×10^{-21}
Ti	2.928×10^{-24}
Al	3.148×10^{-17}

Then the inter-diffusivities can be considered a function of the size fractions of the components in each layer. So far the Fick's Second Law (FSL) could be used to model each of the three layers: Upper Ti, Al and Lower Ti.

Upper Ti Layer:

In the upper Ti layer right next to the Au layer, the partial differential equation (PDE) describing the behavior of the diffusing species governed by the concentration gradient was:

$$\begin{aligned} \frac{\partial C_{Au@Ti1}(x, t)}{\partial t} &= \frac{\partial}{\partial x} \left(\widetilde{D}_1(C_{Au@Ti1}) \frac{\partial C_{Au@Ti1}(x, t)}{\partial x} \right) \\ &= \widetilde{D}_1(C_{Au@Ti1}) \frac{\partial^2 C_{Au@Ti1}(x, t)}{\partial x^2} \end{aligned} \quad (3.26)$$

where $C_{Au@Ti1}(x, t)$ is the concentration of Au atoms in the upper Ti layer. To solve the equation, the initial conditions and the boundary conditions needed to be

specified. Suppose there was an x axis pointing down into the device, with the origin at the top of the Au layer. Since the thickness of the top Au layer was 55 nm, and the thickness of the upper Ti layer was 45 nm, the initial conditions were:

$$C_{Au@Ti1}(x, t = 0) = \begin{cases} 50at.\%Au, & x = 55 \text{ nm} \\ 0, & 55 \text{ nm} < x \leq 100 \text{ nm} \end{cases} \quad (3.27)$$

It was assumed that at the boundary between the Au and the upper Ti layer, the concentration of Au atoms was 50 at.% at $t = 0$, a perfect boundary. Also, there were no Au atoms inside the Ti layer. For the boundary conditions, the boundary between the Au and the upper Ti layer might as well be assumed to maintain a dynamic equilibrium at $50at.\%Au = 50at.\%Ti$, considering the interdiffusion behavior. On the other hand, suppose that the Ti layer was infinitely thick, at $x = \infty$, the concentration would always be 0. Therefore, the boundary conditions were

$$C_{Au@Ti1}(x = 55 \text{ nm}, t) = 50 \text{ at.\% Au} \quad (3.28)$$

$$C_{Au@Ti1}(x = \infty, t) = 0 \quad (3.29)$$

To make the PDE solvable in Mathematica, transformation needed to be made by taking $y = \frac{1}{x}$ so that $x = \infty$ can be transformed to $y = 0$. Therefore, the equation then became

$$\begin{aligned} \frac{\partial C_{Au@Ti1}\left(\frac{1}{y}, t\right)}{\partial t} &= \frac{\partial}{\partial \frac{1}{y}} \left(\frac{\widetilde{D}_1(C_{Au@Ti1}) \partial C_{Au@Ti1}\left(\frac{1}{y}, t\right)}{\partial \frac{1}{y}} \right) \\ &= \frac{dy}{d\frac{1}{y}} \frac{\partial}{\partial y} \left(\frac{d\frac{1}{y}}{dy} \frac{\widetilde{D}_1(C_{Au@Ti1}) \partial C_{Au@Ti1}(y, t)}{\partial y} \right) \\ &= y^2 \frac{\partial}{\partial y} \left(y^2 \frac{\widetilde{D}_1(C_{Au@Ti1}) \partial C_{Au@Ti1}(y, t)}{\partial y} \right) \end{aligned} \quad (3.30)$$

The initial condition became

$$C_{Au@Ti1}(y, t = 0) = 50 \text{ at.}\% Au, y = 1.818 \times 10^7 \text{ m}^{-1}, \quad (3.31)$$

$$1 \times 10^5 \text{ m}^{-1} \leq y < 1.818 \times 10^7 \text{ m}^{-1}$$

The boundary conditions became

$$C_{Au@Ti1}(y = 1.818 \times 10^7 \text{ m}^{-1}, t) = 50 \text{ at.}\% Au \quad (3.32)$$

$$C_{Au@Ti1}(y = 0, t) = 0 \quad (3.33)$$

Al Layer:

Similarly, the PDE of the 100 nm-thick-Al layer was:

$$\frac{\partial C_{Au@Al}(x, t)}{\partial t} = \frac{\partial}{\partial x} \left(\frac{\widetilde{D}_2(C_{Au@Al}) \partial C_{Au@Al}(x, t)}{\partial x} \right) \quad (3.34)$$

where $C_{Au@Al}(x, t)$ is the concentration of Au atoms in the Al layer. Assuming there were no Au atoms throughout the Al layer at $t = 0$. The initial condition was

$$C_{Au@Al}(x, t = 0) = 0 \quad (3.35)$$

The boundary concentration at $x = 100 \text{ nm}$, the upper Ti-Al interface, was dependent on the solutions of the upper Ti layer. Hence, the boundary condition of the Al layer was

$$C_{Au@Al}(x = 100 \text{ nm}, t) = C_{Au@Ti1}(x = 100 \text{ nm}, t) \quad (3.36)$$

$$C_{Au@Al}(x = \infty, t) = 0 \quad (3.37)$$

By taking $y = \frac{1}{x}$, the equation became

$$\frac{\partial C_{Au@Al}(y, t)}{\partial t} = -y^2 \frac{\partial}{\partial y} \left(-y^2 \frac{\widetilde{D}_2(C_{Au@Al}) \partial C_{Au@Al}(y, t)}{\partial y} \right) \quad (3.38)$$

The initial condition became

$$C_{Au@Al}(y, t = 0) = 0 \quad (3.39)$$

The boundary conditions became

$$C_{Au@Al}(y = 1 \times 10^5 \text{ m}^{-1}, t) = C_{Au@Ti1}(y = 1 \times 10^5 \text{ m}^{-1}, t) \quad (3.40)$$

$$C_{Au@Al}(y = 0, t) = 0 \quad (3.41)$$

Lower Ti Layer:

The PDE of the lower Ti layer, which was 20 nm thick, was:

$$\frac{\partial C_{Au@Ti2}(x, t)}{\partial t} = \frac{\partial}{\partial x} \left(\frac{\widetilde{D}_3(C_{Au@Ti2}) \partial C_{Au@Ti2}(x, t)}{\partial x} \right) \quad (3.42)$$

where $C_{Au@Ti2}(x, t)$ was the concentration of Au atoms in the lower Ti layer. The initial condition was

$$C_{Au@Ti2}(x, t = 0) = 0 \quad (3.43)$$

The boundary concentration at $x = 200 \text{ nm}$, the Al-lower Ti interface, was dependent on the solutions of the Al layer. Therefore, the boundary condition of the lower Ti layer was

$$C_{Au@Ti2}(x = 200 \text{ nm}, t) = C_{Au@Al}(x = 200 \text{ nm}, t) \quad (3.44)$$

$$C_{Au@Ti2}(x = \infty, t) = 0 \quad (3.45)$$

By taking $y = \frac{1}{x}$, the equation became

$$\frac{\partial C_{Au@Ti2}(y, t)}{\partial t} = -y^2 \frac{\partial}{\partial y} \left(-y^2 \frac{\widetilde{D}_3(C_{Au@Ti2}) \partial C_{Au@Ti2}(y, t)}{\partial y} \right) \quad (3.46)$$

The initial condition became

$$C_{Au@Ti2}(y, t = 0) = 0 \quad (3.47)$$

The boundary conditions became

$$C_{Au@Ti2}(y = 5 \times 10^4 m^{-1}, t) = C_{Au@Al}(y = 5 \times 10^4 m^{-1}, t) \quad (3.48)$$

$$C_{Au@Ti2}(y = 0, t) = 0 \quad (3.49)$$

The established equations and the initial/boundary conditions were solved in Mathematica. The outcome of the computation, the concentration profile of Au atoms for all the three layers, was shown in Figure 3.5. The profile was then integrated over the Ti/Al/Ti layers to compute the total number of Au atoms in these layers of the source contact, N_{Au} , as a function of the elapsed test time t :

$$N_{Au} = f_d(t) \quad (3.50)$$

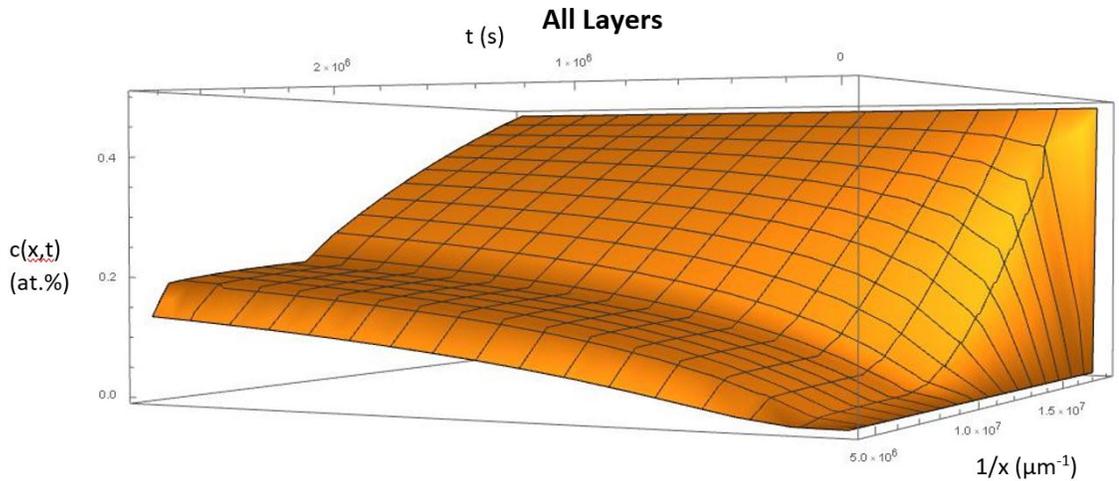


Figure 3.5: Au concentration profile in the Ti/Al/Ti layers of the GaN HEMT source contact vs. time

Consequently, the data series of N_{Au} in the Ti/Al/Ti layers were obtained and taken as the “driving variable” in the system process model of GaN HEMTs. Figure 3.6 shows the plot of N_{Au} over the elapsed test time. The data series of the numbers of Au atoms were converted to the volume (size) of field plate void through the following equation:

$$lv(t) = \frac{M_{Au}N_{Au}}{\rho_{Au}N_A} \quad (3.51)$$

where M_{Au} is the molar mass of Au, ρ_{Au} is the density of Au and N_A is the Avogadro’s constant.

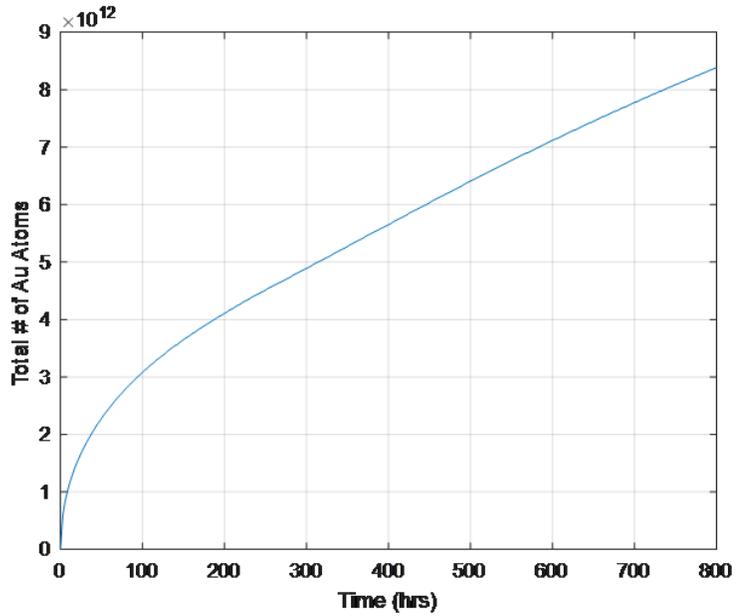


Figure 3.6: The total number of Au atoms in the Ti/Al/Ti layers of the GaN HEMT source contact vs. the elapsed test time

On the other hand, the hypothetical parameter in GaN HEMTs was simply denoted *load*. The associated loading parameters under high-frequency drive conditions were the input power P_{in} , and the junction temperature T_j , as is shown in Equation

3.52. It should be noted that both the loading parameters were kept constant in the HF-ALT, so the load was also constant throughout the test.

$$load = f_l(P_{in}, T_j) \quad (3.52)$$

3.4.3 Model Specification

Once the loading parameter and the “driving variable” were selected and computed, the process model and the measurement model could then be specified.

For IGBT modules, the process model took the following form:

$$s(k+1) = f_{IGBT}(s(k), l(k), \Delta K(k), \omega(k)) \quad (3.53)$$

where $s(k)$ denotes the failure precursor value ($V_{forward}$ for IGBTs) at time step k , and $\omega(k)$ denotes the process noise. Specifying the process model, including $s(k)$ as a function of $l(k)$ and $\Delta K(k)$, as well as $l(k)$ as a function of the loading parameters I_{load} and ΔT_j , required selection of a proper nonlinear mathematical model. The power law polynomial model was selected among a number of common nonlinear mathematical models, including the power law model, the exponential model, multiplicative model, etc., because it had been demonstrated to provide greater flexibility in shape [136], [137]. Corbetta et al. [138] compared and discussed the choice of process noise, with the optimal noise being a multiplicative noise where a lognormal random noise e^ω , $\omega \sim N(0, \sigma_\omega^2)$ was multiplied by the increments to represent the uncertainties. The ω in the noise followed the normal distribution and was constrained by the principle below, as was validated in [138]:

$$\mu_\omega = -\frac{\sigma_\omega^2}{2} \quad (3.54)$$

Therefore, the process model of IGBT modules related to the initial value of the failure precursor s_0 is shown in Equation 3.55:

$$s(k) = s_0 + (l(k) - l_0 + m_1 \Delta K(k)^{n_1}) e^{\omega(k)} \quad (3.55)$$

The load variable $l(k)$ was also assumed to follow the power law polynomial model:

$$l(k) = m_2 I_{load}^{n_2} + m_3 \Delta T_j(k)^{n_3} \quad (3.56)$$

Taking the derivative of Equation 3.55 and combining Equation 3.56 leads to:

$$\frac{ds(k)}{dt} = \left(m_3 n_3 \Delta T_j(k)^{n_3-1} \frac{\Delta T_j(k)}{\partial t} + m_1 n_1 \Delta K(k)^{n_1-1} \frac{\Delta K(k)}{\partial t} \right) e^{\omega(k)} \quad (3.57)$$

By assuming that dt can be discretized into small steps, Equation 3.57 was converted to the following canonical form:

$$\begin{aligned} s(k+1) = s(k) &+ (m_3 n_3 \Delta T_j(k)^{n_3-1} (\Delta T_j(k+1) - \Delta T_j(k))) e^{\omega(k)} \\ &+ (m_1 n_1 \Delta K(k)^{n_1-1} (\Delta K(k+1) - \Delta K(k))) e^{\omega(k)} \end{aligned} \quad (3.58)$$

whereas

$$s(1) = s_0 + (l(1) - l_0 + m_1 \Delta K(1)^{n_1}) e^{\omega(1)} \quad (3.59)$$

The measurement model described the functional relationship between the measured values and the failure precursor. In the present work, since the failure precursor was directly measurable, the measurement model only accounted for the differences between the actual measurement and the actual failure precursor value, i.e. the measurement error. The measurement model can be depicted as a de-noising step. For IGBT modules, the two variables measured in-situ and monitored throughout the test are $V_{forward}$ and ΔT_j , and they were de-noised respectively. The measurement

error was assumed to be Gaussian, and therefore the measurement model is as follows:

$$y(k) = s(k) + v_1(k) \quad (3.60)$$

$$\Delta T_y(k) = \Delta T_j(k) + v_2(k) \quad (3.61)$$

where $y(k)$ and $\Delta T_y(k)$ are the raw measurements of $s(k)$ and $\Delta T_j(k)$, and $v_1(k)$ and $v_2(k)$ denote the measurement noises of $s(k)$ and $\Delta T_j(k)$, respectively. In summary, the complete prognostic model for IGBT modules under power cycling conditions can be assembled and shown below:

$$\begin{aligned} (PC_i, SV_i) &= f_{PCA} \left(\frac{\bar{d}s}{dt}, \sigma \frac{dx_{ij}}{dt}, CF \frac{dx_{ij}}{dt}, \frac{ds}{dt}_{max}, \frac{ds}{dt}_{rms}, IF \frac{ds}{dt}, SF \frac{ds}{dt}, ED \frac{ds}{dt}, CLF \frac{ds}{dt} \right) \\ d_{j,m}(k) &= \left[\left(\frac{SV_1 (PC_1(k) - PC_{1,m}^j)}{SV_1 + SV_2 + \dots + SV_n} \right)^2 + \left(\frac{SV_2 (PC_2(k) - PC_{2,m}^j)}{SV_1 + SV_2 + \dots + SV_n} \right)^2 \right. \\ &\quad \left. + \dots + \left(\frac{SV_n (PC_n(k) - PC_{n,m}^j)}{SV_1 + SV_2 + \dots + SV_n} \right)^2 \right]^{1/2} \\ P_{det,j}(k) &= f_{det}(d_{j,m}(k), d_{j,m_1,m_2}(k)) \end{aligned} \quad (3.62)$$

$$s(k) = s(k-1) + (m_3 n_3 \Delta T_j(k-1))^{n_3-1} (\Delta T_j(k) - \Delta T_j(k-1))$$

$$+ m_1 n_1 \Delta K(k-1)^{n_1-1} (\Delta K(k) - \Delta K(k-1)) e^{\omega(k)}$$

$$l(k) = m_2 I_{load}^{n_2} + m_3 \Delta T_j(k)^{n_3}$$

$$y(k) = s(k) + v_1(k)$$

$$\Delta T_y(k) = \Delta T_j(k) + v_2(k)$$

$$[s(0) \quad \Delta T_j(0) \quad I_{load}] = [s_0 \quad \Delta T_{j0} \quad I_{load}]$$

where the first three equations constitute the anomaly detection model, $f_{PCA}()$

denotes the feature extraction process using PCA, while the 9 independent variables inside denote the 9 time-domain statistical features constructed for feature extraction. PC_i and SV_i denote the outcome of PCA, the i th principal component and the corresponding singular value, respectively. $i = 1, 2, \dots, n$ with n being the number of PCs. In the second equation, $d_{j,m}$ denotes the SVWD between the current test data instance and the m th healthy data instance in the j th baseline cluster. $j = 1, 2, \dots, l$ with l being the number of baseline clusters, and $m = 1, 2, \dots, h$ with h being the number of healthy data instances in the j th baseline cluster. $PC_{i,m}^j$ denotes the coordinate in the i th dimension in the feature space constructed by PCs, of the m th healthy data instance inside the j th cluster. $f_{det}()$ denotes the outlier probability computation process based on FC. $P_{det,j}$ denotes the outlier probability for the j th baseline cluster, while d_{j,m_1,m_2} is the SVWD between the m_1 th and the m_2 th healthy data instances within the j th healthy cluster. Next to the anomaly detection model are the process model and the measurement model. The last row describes the initial conditions of the failure precursor s , denoted by s_0 , and the loading parameters ΔT_j and I_{load} , respectively.

For GaN HEMTs, on the other hand, the process model takes the following form:

$$s(t + \Delta t) = f_{GaN}(s(t), load, lv(t), \omega(t)) \quad (3.63)$$

where $s(t)$ denotes the value of P_{out} , the failure precursor of GaN HEMTs at time t , and $\omega(t)$ denotes the process noise at time t . The process model, including $s(t)$ as a function of $load$ and $lv(t)$, as well as $load$ as a function of the loading parameters

I_{load} and ΔT_j , is also specified using the power law polynomial model and the optimal noise used for IGBT modules. Therefore, the process model of GaN HEMTs related to the initial value of the failure precursor s_0 is shown below:

$$s(t) = s_0 + (m_1 * load * lv(t)^{n_1}) e^{\omega(t)} \quad (3.64)$$

The hypothetic loading parameter as a function of I_{load} and T_j :

$$load = m_2 P_{in}^{n_2} + m_3 T_j^{n_3} \quad (3.65)$$

Taking the derivative of Equation 3.64 gives the stepwise evolution model of $s(t)$:

$$\frac{\partial s(t)}{\partial t} = m_1 n_1 * load * lv(t)^{n_1-1} \frac{\partial lv(t)}{\partial t} e^{\omega(t)} + (m_1 * load * lv(t)^{n_1}) e^{\omega(t)} \frac{\partial \omega(t)}{\partial t} \quad (3.66)$$

Assuming that ∂t can be discretized into steps with a Δt gap, Equation 3.66 was converted to the following form:

$$\begin{aligned} s(t) = & s(t - \Delta t) + m_1 n_1 * load * lv(t - \Delta t)^{n_1-1} (lv(t) - lv(t - \Delta t)) e^{\omega(t)} \\ & + (m_1 * load * lv(t - \Delta t)^{n_1}) e^{\omega(t-\Delta t)} (\omega(t) - \omega(t - \Delta t)) \end{aligned} \quad (3.67)$$

whereas

$$s(\Delta t) = s_0 + (load + m_1 lv(\Delta t)^{n_1}) e^{\omega(\Delta t)} \quad (3.68)$$

For GaN HEMTs, only one variable, P_{out} , was monitored in-situ throughout the test. The measurement error was also assumed to be Gaussian, and the measurement model becomes:

$$P_{out,y}(t) = P_{out}(t) + v(t) \quad (3.69)$$

where $P_{out,y}(t)$ denotes the raw measurement at t , and $P_{out}(t)$ denotes the measurement noise of P_{out} at t . In summary, the complete prognostic model for GaN

HEMTs under RF-ALT conditions could be assembled and shown below:

$$\begin{aligned}
(PC_i, SV_i) &= f_{PCA} \left(\frac{\bar{d}s}{dt}, \sigma_{\frac{dx_y}{dt}}, CF_{\frac{dx_y}{dt}}, \frac{ds}{dt}_{max}, \frac{ds}{dt}_{rms}, IF_{\frac{ds}{dt}}, SF_{\frac{ds}{dt}}, ED_{\frac{ds}{dt}}, CLF_{\frac{ds}{dt}} \right) \\
d_{j,m}(t) &= \left[\left(\frac{SV_1 (PC_1(t) - PC_{1,m^j})}{SV_1 + SV_2 + \dots + SV_n} \right)^2 + \left(\frac{SV_2 (PC_2(t) - PC_{2,m^j})}{SV_1 + SV_2 + \dots + SV_n} \right)^2 \right. \\
&\quad \left. + \dots + \left(\frac{SV_n (PC_n(t) - PC_{n,m^j})}{SV_1 + SV_2 + \dots + SV_n} \right)^2 \right]^{1/2} \\
P_{det,j}(t) &= f_{det}(d_{j,m}(t), d_{j,m_1,m_2}(t)) \tag{3.70}
\end{aligned}$$

$$\begin{aligned}
s(t) &= s(t - \Delta t) + m_1 n_1 * load * lv(t - \Delta t)^{n_1 - 1} (lv(t) - lv(t - \Delta t)) e^{\omega(t)} \\
&\quad + (m_1 * load * lv(t - \Delta t)^{n_1}) e^{\omega(t - \Delta t)} (\omega(t) - \omega(t - \Delta t))
\end{aligned}$$

$$load = m_2 P_{in}^{n_2} + m_3 T_j^{n_3}$$

$$P_{out,y}(t) = P_{out}(t) + v(t)$$

$$s(0) = s_0$$

The framework of the model was kept the same as IGBT modules, with the last row also concerning the initial conditions. Since the input loading conditions remained the same throughout the test, only the initial value of failure precursor needed to be specified.

3.5 Summary

This chapter presented in detail the development of the prognostic model including a novel anomaly detection module and a RUL prediction module including a non-linear process model and a measurement model. The anomaly detection model utilized semi-supervised machine learning, with the healthy baseline established based

on unsupervised machine learning techniques, including PCA for dimensionality reduction and k-means clustering for partitioning the baseline healthy data in the "nominal region". The determination of the outlier probability of a test data instance for each baseline cluster was based on calculating the Fisher criterion of the distribution of distances between data instances within each baseline cluster, and the distribution of distances between the test data instance and the healthy data instances within the baseline cluster. The developed anomaly detection model was validated on the famous wine dataset evaluating the performance of clustering algorithms and outperformed the Mahalanobis-distance-based approach. The RUL prediction module included a power-law polynomial process model and a measurement model. The process model included a "driving variable" and a hypothetical loading variable, which were tailored to the physics-of-failure and the type of stress the specific power devices were subject to. For IGBT modules, the "driving variable" was the stress intensity factor whose relationship with the elapsed power cycles was derived from Paris equation. For GaN HEMTs, the "driving variable" was the void size in the field plate as a result of ohmic contact interdiffusion, which was obtained from a diffusion equation computation based on temperature profile simulation in COMSOL Multiphysics and equation solving in Mathematica. The SIR filter including 100 particles was implemented for RUL prediction 10 times for each IGBT module, obtaining the average predicted mean CTF and the 95% confidence interval. For GaN HEMTs, since the specified failure threshold was not reached, the average predicted state and the 95% confidence interval at the end of test were also computed 10 times. Overall, the average errors for all the IGBT module and GaN HEMT

degradation data series spanned from 0 to 7%, which were quite satisfactory.

Chapter 4: Novel Resampling Technique Development

4.1 Introduction of Resampling

The prognostic model introduced in the previous chapter aims at providing a timely warning of degradation and potential failure and a robust RUL prediction. The robustness of RUL prediction is evaluated by two criteria. One is the accuracy of prediction, determined by the mean error of prediction. The other is the efficiency of prediction, which for PF is determined by the effectiveness of particles. In other words, in an efficient particle filter, at least the majority of the particles should be effective and play an infusive role in the final RUL distributions. The metric for the efficiency of prediction is the effective number of particles. This effectiveness is influenced by variances in particle distributions. Larger variances reduce the effectiveness of particles, resulting in particle degeneracy, which can be improved by resampling.

The key idea of the particle filter technique is to represent the posterior probability density function (PDF) with a large number of particles and their associated weights. Resampling is an important method tackling the issue of particle degeneracy. At each time step, an original set of particles and their weights are obtained formulating a discretized description of the posterior distribution. Conventionally,

the resampling step is triggered when N_{eff} falls below a pre-defined threshold, compulsively maintaining an acceptable N_{eff} as time elapses. A new particle set the same size as the original set is derived to replace the original set. In this process, particles with low weights are removed and particles with high weights are intensified (often through replication) to increase the effective particle size N_{eff} [139]. More new particles are generated in the region of particles with large weights [9], hereby improving the quality of posterior distribution estimation.

Conventional resampling techniques, while solving the degeneracy issue, could introduce new undesirable effects. The most notable issue is sample impoverishment, i.e. loss of diversity in the newly sampled particles because those particles with high weights in the original set will be duplicated many times in the resampling process, replacing the low-weight particles [139], [140], [141]. Typically, under a severe particle degeneracy where resampling is required, the weights are concentrated on a small portion of all the particles. In such cases, these high-weight particles have much higher probability of being selected in the resampling process, resulting in a small number of unique particles in the new set. Other possibly undesirable effects, such as slower speed of computation due to introduction of resampling, comes down to parallelizing the PF, which is challenging to realize and is not within the scope of the present dissertation.

To eliminate these undesired effects, advanced resampling techniques have been investigated. Common ideas include varying the number of particles, removing the low-weight particles, not resetting all the weights to uniformity, etc. [9]. Resampling techniques can also be classified by the number of distributions particles are resampled

from, single or multiple (compound sampling). For the latter, particles are partitioned into groups before resampling is implemented. The following sections will discuss the problem of traditional resampling techniques, and introduce the novel MHA-replacement resampling technique developed in the present work.

4.2 Traditional Resampling Techniques

Traditional resampling techniques discussed here include multinomial resampling, systematic resampling, residual resampling and stratified resampling. They are jointly called “the classic four”. These four techniques are very similar in that they all resample based on multinomial distribution of particle weights. Let N be the size of the particle set, selecting N particles \tilde{x}_j from the set following the multinomial weight distribution to constitute a new particle set $P(\tilde{x}_j = x_i) = w_i$, $i, j = 1, \dots, N$, and replace the original set. For the new particles, their weights are reset to $1/N$, or uniform weights. The resampling step is triggered when the effective number of particles drop below a threshold, often $2/3$ of N . The differences between the multinomial, systematic and stratified resampling lie in how the random numbers between 0 and 1 corresponding to the cumulative sum of weight in the discretized posterior distributions are generated.

4.2.1 Multinomial Resampling

Multinomial resampling independently generates N random numbers following the uniform distribution between 0 and 1, $U(0, 1)$. Then for each random number,

the particle is selected where the cumulative sum of weights before counting in the particle weight is less than the random number, while after counting in the particle weight the cumulative sum of weights becomes larger than the random number. Since the sampling of each particle is random, the number of times a given particle is resampled ranges from 0 to N .

4.2.2 Systematic Resampling

Systematic resampling is another technique resampling from multinomial distributions. Unlike multinomial resampling, only one random number u_1 is drawn from the uniform distribution $U(0, 1)$, and the other $N - 1$ u numbers are obtained deterministically from $u_i = \frac{u_1 + (i-1)}{N}$, where $i = 2, \dots, N$. Then the same process of particle selection as multinomial resampling is performed. Systematic resampling is considered computationally efficient as it minimizes the number of sampling required.

4.2.3 Stratified Resampling

Stratified resampling divides the $(0, 1)$ interval into N equal sub-intervals. In each sub-interval, one sample is generated following uniform distribution within this sub-interval and the same process of particle selection as multinomial resampling is performed again. Figure 4.1 gives the pseudo-code of multinomial, systematic and stratified resampling.

4.2.4 Residual Resampling

Residual resampling is slightly different from the three techniques introduced above. However, the idea is still resampling following the multinomial distribution of weights. There are two stages in residual resampling. First, the majority of N particles are resampled deterministically by capturing $n_i = \lfloor Nw_i \rfloor$ copies of the i th particle in the original set. Obviously, for those particles with weights below $1/N$, no copy is captured in the first stage. $r_i = Nw_i - n_i$ is called the residual of the i th particle. At the second stage, the remaining particles to be selected, which is the difference between N and total number of samples already generated, are sampled based on multinomial sampling proportional to the weights of residuals. Figure 4.2 gives the pseudo-code of residual resampling.

```

 $\{\hat{x}_t^i\}_{i=1}^N = \text{Resample}[\{x_t^j, w_t^j\}_{j=1}^N, N]$ 
 $\{Q_t^i\}_{i=1}^N = \text{CumulativeSum}[\{w_t^j\}_{j=1}^N]$ 
i=0
/systematic/stratified runs:
  j=1
/systematic runs:
   $u_1 \sim U(0,1)$ 
while (i ≤ N)
/stratified runs:
   $u_1 \sim U(0,1)$ 
/systematic/stratified runs:
   $u_i = u_1 + \frac{i-1}{N}$ 
/multinomial runs:
   $u_i \sim U(0,1); j = 1$ 
while ( $Q_t^j < u$ )
  j=j+1
end
  i=i+1
   $\hat{x}_t^i = x_t^j$ 
end

```

Figure 4.1: Pseudo-code of multinomial, systematic and stratified resampling [9]

```

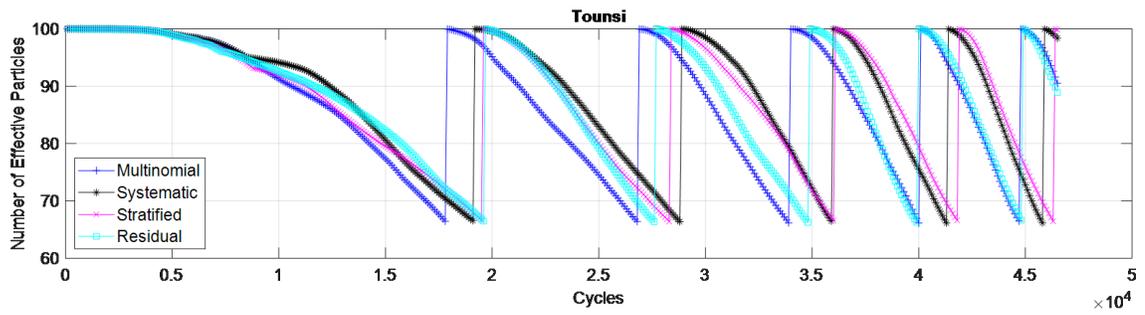
 $\{\tilde{x}_t^i\}_{i=1}^N = \text{Resample}[\{x_t^j, w_t^j\}_{j=1}^N, N]$ 
for j=1:N
     $N_t^j = \text{floor}(N \times w_t^j)$ 
     $\widehat{w}_t^j = w_t^j \cdot \frac{N}{N_t^j}$ 
end
 $[\{\tilde{x}_t^i\}_{i=1}^N, N_t] = \text{Replication}[\{x_t^j, N_t^j\}_{j=1}^N]$ 
for j=1:N
     $\widehat{w}_t^j = \widehat{w}_t^j \times \frac{N}{N - N_t}$ 
end
 $\{\tilde{x}_t^i\}_{i=N_t+1}^N = \text{Resample}[\{x_t^j, \widehat{w}_t^j\}_{j=1}^N, N - N_t]$ 

```

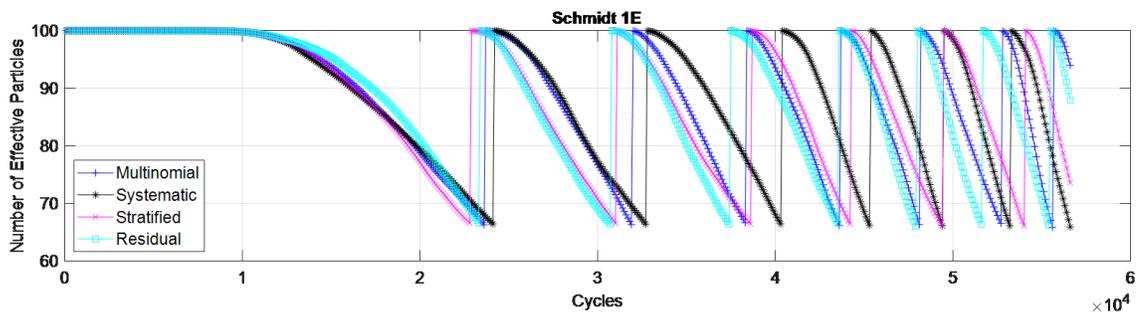
Figure 4.2: Pseudo-code of residual resampling [9]

The four traditional resampling techniques introduced above are well-known and extensively used. There are a few variants of these traditional techniques, such as the residual systematic resampling (RSR) [142], branching [143], and rounding-copy resampling [144], all of which are not essentially different from “the classic four”. Overall, these techniques do alleviate the degeneracy problem by performing resampling whenever the effective sample size N_{eff} drops below a threshold, and resetting all the weights in the new particle set to uniformity ($1/N$). Nevertheless, the drawbacks of these traditional techniques, other than sample impoverishment are rarely reviewed and discussed in the literature. On the other hand, sample impoverishment can be severe in these techniques, and the severity actually depends on the threshold set for N_{eff} . If a very low N_{eff} threshold is set, severe degeneracy occurs, so the weights concentrate on a small number of particles, leading to repeated selection of these particles in traditional resampling techniques. This results in severe sample impoverishment. If the N_{eff} threshold is not set too low, then resampling

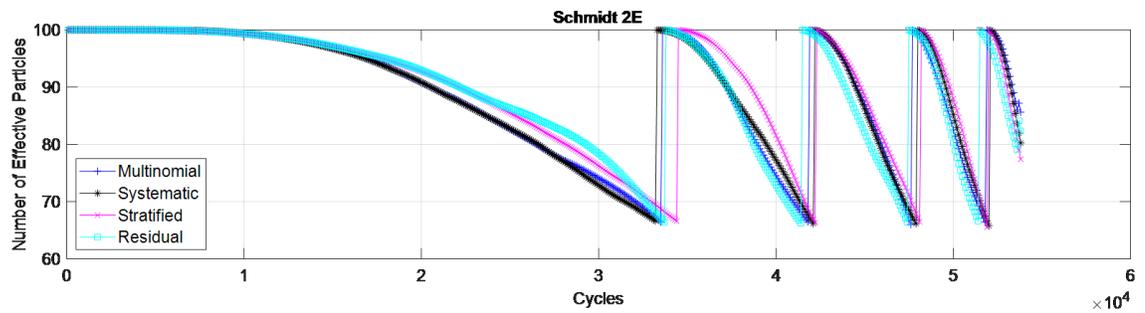
will occur more frequently, raising another problem as will be discussed next.



(a) Tounsi



(b) Schmidt 1E



(c) Schmidt 2E

Figure 4.3: Effective sample size N_{eff} vs. cycles for "the classic four" techniques in long-term dynamic state estimation showing increasingly frequent resampling and weight reset

The real problem of "the classic four" techniques is the increasing frequency of resampling as time elapses in Bayesian tracking problems. Figure 4.3 shows the

variation of the effective sample size N_{eff} with time when implementing “the classic four” techniques on three sets of IGBT module degradation data studied in the present dissertation. It is obvious that as time (number of power cycles) elapsed, the gap between adjacent resampling became shorter and shorter. i.e. The decrease in N_{eff} speeded up over time, hitting the threshold in a shorter time period. The reason behind this phenomenon is increased scattering of particles after accumulated recursions of Bayesian filtering process due to its Markov property, even with the resampling step. This increasing frequency of resampling can be a significant issue for long-term dynamic tracking and estimation. Every time resampling is executed, the weights of the resampled particle set is reset to uniformity, which means loss of previous weight information. Since weight update is only based on the weight information at the previous time step, with weight resetting the previous weight information cannot be recovered. Therefore, as an anomaly is detected, it becomes increasingly likely that the weights are just set to uniformity, making the RUL prediction utilize uniform or close to uniform weights. This makes no sense for particle filters because the essence of particle filters is utilizing discriminated set of weights to estimate the most likely future states. Uniform weights create barriers for inference of the most likely estimates of the state. It may also increase the sensitivity of the posterior state estimation to outliers. All in all, traditional resampling allows high-weight particles to be more pronounced in the particle set, but at the price of sample impoverishment. The accompanying frequent reset of weights to uniformity is likely deteriorating the quality of posterior state distribution estimation.

4.3 Other Resampling Techniques

In addition to the aforementioned traditional techniques, more advanced resampling techniques have been developed. For examples, in contrast to traditional techniques sampling from a single distribution for all particles, some advanced techniques feature partitioning the particles into groups before resampling. To speed up the computation, parallel resampling has been developed for complex problems conducting resampling tasks for multiple particles simultaneously, and grouping of particles is necessary under these circumstances. In the present work, Gaussian process noise was assumed for the process model in the propagation (future state estimation) of particles. At each time step the distribution of particle positions is expected to follow Gaussian distribution. There is no need of grouping particles or resampling from multiple distributions. Therefore, the techniques discussed in the present work still focus on single-distribution resampling. Detailed review of compound resampling can be found in [9].

There have been several other resampling techniques proposed. Modified resampling [145] provides a more generalized and flexible framework based on conventional multinomial resampling through introducing a power law parameter α to the weights. i.e. The distribution of weights from which particles are drawn becomes the distribution of a power-law function:

$$p_t^i \propto (w_t^i)^\alpha \quad (4.1)$$

where p_t^i is the probability of the i th particle being resampled, w_t^i denotes the weight

of the i th particle. When $0 < \alpha < 1$, the influences of lower weight particles get intensified, while those of higher weight particles get attenuated, and vice versa for $\alpha > 1$. The challenge of implementing modified resampling is the choice of α . Practically, it adds significant complexity to the algorithm entailing analytical assessment of a proper value of α based on the set of weights at each time step.

Also, modified resampling is still resampling from the original set of weights, subject to either severe sample impoverishment with larger α or perturbation of low-weight particles in posterior distribution formulation with smaller α . Variable-size resampling [146] is another technique introducing flexibility into multinomial resampling featuring determination of the required sample size at every time step. The optimal number of particles is determined so that the Kullback–Leibler divergence (KLD) between the sample-based maximum likelihood estimate (MLE) of a desired distribution and the actual distribution is less than a predefined error bound ε , with the probability of $1 - \rho$:

$$N = \frac{1}{2\varepsilon}q \tag{4.2}$$

where

$$q = F^{-1}(1 - \rho) \tag{4.3}$$

where $F^{-1}()$ is the inverse of the cumulative chi-squared distribution with $k-1$ degrees of freedom, k being the number of non-overlapping multidimensional intervals to sort the particles. The problem of this technique is also the complexity of computation and again, reliance on multinomial resampling as the basis. Both of the techniques are essentially tweaking the traditional techniques, while inheriting their potential

problems.

A resampling technique distinct from all the previously discussed resampling techniques is independent Metropolis-Hastings algorithm (IMHA) [9], [10], [147]. In Metropolis-Hastings algorithm (MHA), particles are generated from the proposal distribution $q(\tilde{x}_t^i|x_t^i)$, which is the probability of returning a resampled particle of \tilde{x}_t^i given a particle x_t^i in the original set. Then an acceptance probability is computed based on the conditional density function and independent proposal distributions for both \tilde{x}_t^i and x_t^i . IMHA is a special case of MHA replacing the conditional density functions with independent density functions. The pseudo-code of IMHA is shown in Figure 4.4 below:

```

Initialize  $x_0 \sim q(x)$ 
for particle  $i = 1, 2, \dots, N$  at time step  $t$ 
  Propose:  $x^* \sim q(x_t^i|x_{t-1}^i)$ 
  Acceptance Probability:  $\alpha(x^*|x_{t-1}^i) = \min\{1, \frac{q(x_{t-1}^i)\pi(x^*)}{q(x^*)\pi(x_{t-1}^i)}\}$ 
   $u \sim U(0,1)$ 
  if  $u < \alpha$ 
    Accept the proposal:  $x_t^i \leftarrow x^*$ 
  else
    Reject the proposal:  $x_t^i \leftarrow x_{t-1}^i$ 
  end
  Normalize the weights:  $\omega_k^i = \frac{\omega_k^i}{\sum_{i=1}^N \omega_k^i}$ 
end

```

Figure 4.4: Pseudo-code of IMHA resampling [10]

IMHA resampling, unlike the traditional resampling techniques and their variants discussed earlier, does not produce the new particle set based on the multinomial weight distribution of the original particle set. It is generating a new candidate particle at each iteration from the proposal distribution straightly and let

it compete with the available particle based on the acceptance probability. There are different versions of IMHA resampling in dealing with weight computation. One version involves resetting the weights to $1/N$ [9], while the other version performs weight normalization given that upon acceptance, both the new particle and its associated weight replace the original particle and its associated weight [10], [147]. The latter version gets around the sample impoverishment issue as there is little chance of selecting identical particles multiple times. To avoid the drawback of resetting weights to uniformity discussed earlier, the present work uses the version with weight normalization for benchmarking purposes. Degeneracy is probably still an issue in IMHA due to elimination of weight reset to uniformity. The severity is expected not as great as the SIS particle filter without resampling, but it is still going to be a concern.

Another resampling technique that essentially introduces an additional step to the conventional resampling techniques to counteract the sample impoverishment issue is called roughening resampling [100]. The basic idea of this technique is quite simple: since in conventional techniques like multinomial resampling, sample impoverishment occurs due to repeatedly selecting particles with high weights, addition of a random noise called roughening noise to the resampled particles could help decentralize or roughen their values and therefore rejuvenate the diversity of particles. A Gaussian noise with zero mean and constant variance is commonly selected for the roughening noise. The variance can be selected using the following equation [100]:

$$\sigma = KDN^{-\frac{1}{d_x}} \quad (4.4)$$

where K is a user-selected positive constant, D is the difference between the maximum and minimum values of the state variable, N is the number of particles, and d_x denotes the number of dimensions of the state variable. One technique with the similar idea is called the resample-move algorithm. This technique adds a move step after the resampling step based on Markov chain Monte Carlo (MCMC) sampling method to rejuvenate the diversity of particles. The move step performs one or more iterations on each of the resampled particles to sample its position at the next time step, and then “move” the resampled particle to the sampled position. The sampling guiding the “move” is based on a transition kernel with the posterior PDF at the next step as the invariant distribution [148]. All in all, roughening resampling can theoretically both resolve degeneracy issue and alleviate sample impoverishment through the roughening step. Therefore, it was considered a promising candidate of a robust resampling technique and selected for benchmarking in the present dissertation.

4.4 Novel Resampling Technique: IMHA-Replacement Resampling

Review of the state-of-the-art resampling techniques showed that the majority of them perform reset of weights to uniformity, given the effective sample size N_{eff} reaches a threshold. Despite instant recovery of N_{eff} , there are severe limitations of this approach for long-term Bayesian tracking problems. Therefore, a novel resampling technique was developed in the present dissertation avoiding resetting the weights to uniformity. The technique also resolves the degeneracy issue by intentionally and selectively replacing the low-weight particles ($weight < 1/N$) with

new particles sampled from the high-weight region, where particle weight is greater than $1/N$. This new resampling technique is called IMHA-replacement resampling. In the present dissertation, the resampling step was implemented at every time step in the particle-filter-based MCMC sampling process, without setting a threshold of N_{eff} to trigger resampling. Despite likely increase in computation time, drastically eliminating degeneracy while avoiding severe sample impoverishment is the major concern in the present work.

IMHA-Replacement resampling consists of two steps. The first step is the conventional IMHA resampling, which improves the effective sample size slightly without introducing sample impoverishment. The second step is called replacement step, which captures all the particles with below-average weights ($weight < 1/N$) at time step k , assuming the number is N_{sub} and substitute for them using new high-weight particles. The set of new particles used for substitution are selected from a pool of samples twice its size, $2N_{sub}$. Each sample in this pool is resampled from a high-weight particle at the last time step $k - 1$, whose propagation based on the transition probability $p(i, j) = Pr(x_k = j | x_{k-1} = i)$ generates a new particle with higher weight at k , in the importance sampling process before resampling. i.e. The set of particles with above-average weights in the importance sampling are traced one step back for their positions at $k - 1$. Then an invariant distribution is formed with the high-weight set at $k - 1$ to perform importance sampling, selecting a particle and propagate it from $k - 1$ to k . A new particle is therefore attained and used to substitute an original low-weight particle. This resampling process occurs recursively until all the low-weight particles have been substituted. Finally, the

weights of the resampled particle set are normalized. Propagation from the prior positions of high-weight particles at the current step can significantly increase the chance the resampled particles is located in the high-weight region. Simultaneously, this process detours sampling original high-weight particles multiple times as in traditional resampling techniques, keeping the sample impoverishment at a very low level. Selecting half of the particles with higher weights further suppresses the occurrence of particle degeneracy. Figure 4.5 shows the flow chart of the IMHA-replacement resampling technique. Validation of this novel resampling method will be shown in the subsequent chapters.

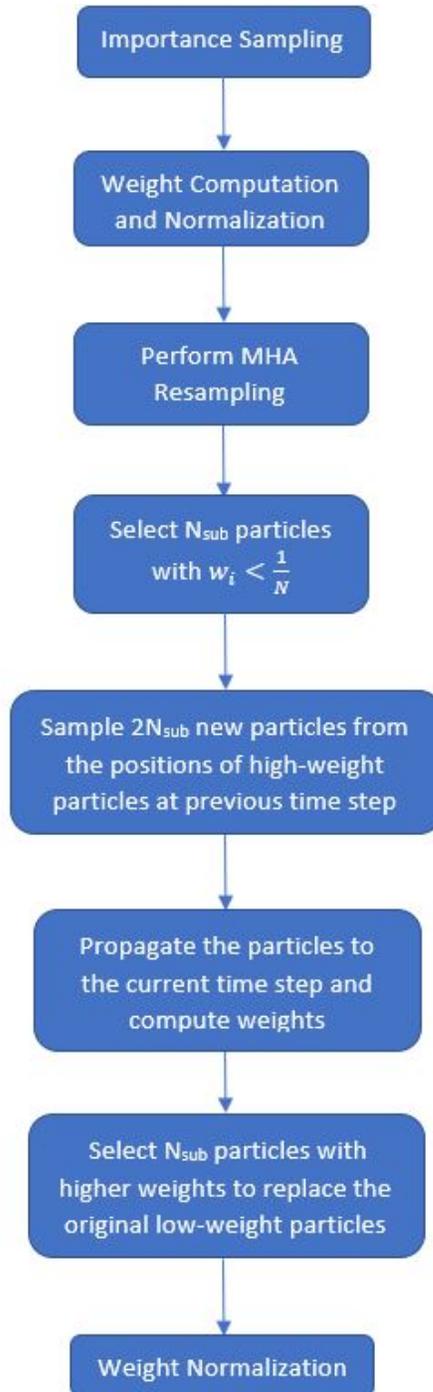


Figure 4.5: Flow chart of the IMHA-replacement resampling technique

4.5 Summary

This chapter reviewed the state-of-the-art resampling techniques, including traditional techniques “the classic four”: multinomial, systematic, stratified and residual resampling, and their variants, including modified and variable-size resampling. All these resampling techniques have an essential problem of resampling multiple copies of high-weight particles in the original particle set. Also, the resampling step is implemented only if the effective sample size N_{eff} decrease below a critical threshold, and the weights are reset to uniformity in resampling. Such process produces negative consequences in long-term prediction of dynamic Bayesian estimation problems. Therefore, weight reset to uniformity and resampling based on N_{eff} should be avoided. Some more advanced techniques developed by researchers that detour the problems above are also discussed and selected for benchmarking, including IMHA resampling and roughening resampling. A novel resampling technique, called IMHA-replacement resampling, is developed. The idea is to replace the low-weight particles with high-weight particles resampled and selected based on the positions of high-weight particles at the previous time step.

Chapter 5: Validation of Prognostic Model for IGBT Modules

5.1 Degradation Data Sources

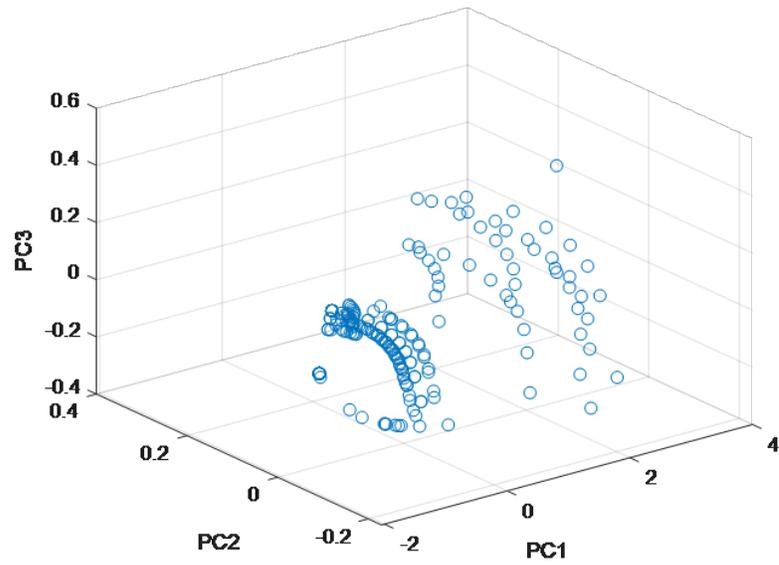
Chapter 3 discussed the development of the prognostic model integrating anomaly detection and RUL prediction, while Chapter 4 discussed the development of the novel IMHA-replacement resampling technique. This chapter will present the validation of prognostic model and the resampling technique on multiple sets of time-series degradation data of IGBT modules. These degradation data series were collected from three different open literature sources reporting IGBT power cycling test results [8], [11], [12]. Power cycling tests were conducted in these sources on commercial 1200 V IGBT modules from different manufacturers with current ratings varying from 15 A to 313 A, under junction temperature swings ΔT_j varying from 70 K to 108 K and different levels of junction temperature extremes $T_{j,min}$ and $T_{j,max}$ as well. Table 5.1 summarizes the power ratings and important test parameters of the IGBT module power cycling degradation data series used in the present dissertation.

Table 5.1: Optimal number of clusters and maximum silhouette values of IGBT module degradation data series

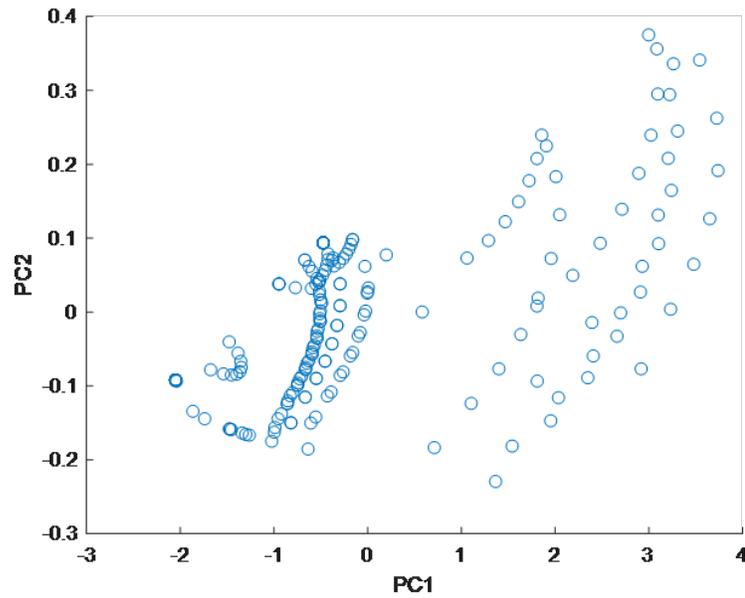
Literature Source	Voltage Ratings V_{CES} (V)	Current Ratings I_C (25 °C)	Junction Temperature Swing ΔT_j (K)	Minimum/Maximum Junction Temperature $T_{j,min}/T_{j,max}$ (°C)
Tounsi (2010)	1200	313	70	80/150
Scheuermann (2011)	1200	15	70	90/160
Schmidt (2013)	1200	300	108	-20/88

5.2 Results of Anomaly Detection

The failure precursor of IGBT modules is the on-state forward voltage drop $V_{forward}$, while the failure threshold is 5% increase in $V_{forward}$. The “nominal region” containing healthy data was defined using the range discussed in Section 3.3. A moving window containing 10 most recent data instances within the “nominal region” was established to compute the statistical features. Then feature extraction based on PCA was performed to get a set of PCs and establish feature space based on these PCs, and the healthy data instances were located inside the space. Figure 5.1 shows the 3D and 2D plots of the healthy $V_{forward}$ data instances from [12] in the PC-based feature space, whose coordinate spaces are formed by the three most important PCs (PC_1 to PC_3) and the two most important PCs (PC_1 and PC_2) corresponding to the largest singular values, respectively.



(a)



(b)

Figure 5.1: Plots of healthy data instances in the PC-based feature space including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only

Then the conventional k -means clustering was implemented to establish the

baseline clusters for these healthy data. The silhouette values with k ranging from 2 to 6 were observed. For the selection of k , Figure 5.2 shows the plot of silhouette value vs. k for the time-series IGBT module degradation data from [8]. Table 5.2 summarizes the values of k for all the degradation data series of IGBT modules studied in the present work. It can be seen that k varied from 2 to 4. Figure 5.3 shows the 3D and 2D plots of the healthy data instances from [12] partitioned into k clusters in the feature space. The points in the same shape and the same color belong to the same cluster. Since the PCs were weighed in the SVWD, the clustering process heavily relied on variance of the healthy data along the axis of PC_1 , as is evident in the figure. This is because PC_1 often takes up the majority (over 80-90%) of the weights because it accounts for most of the variance in the data.

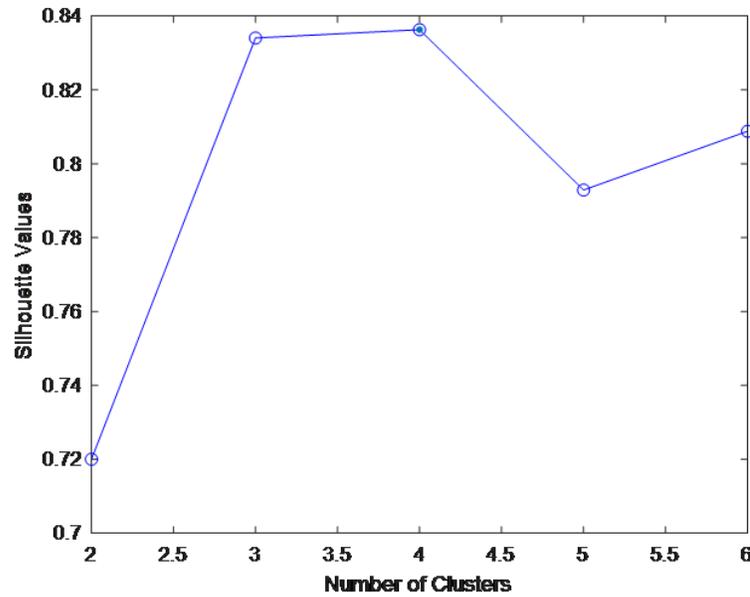
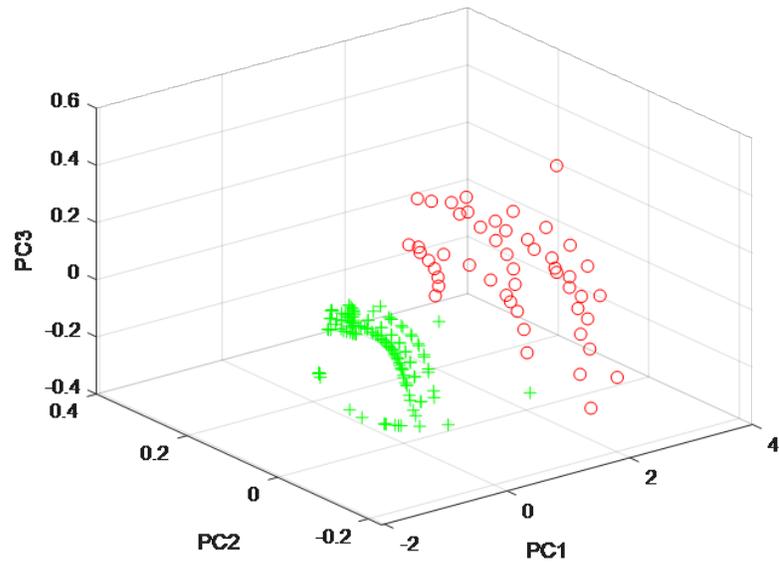
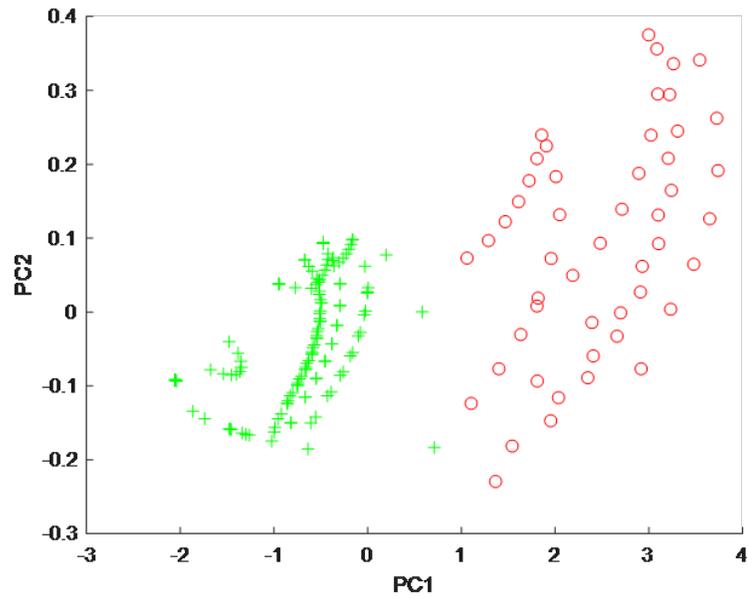


Figure 5.2: Variation of silhouette value with the number of clusters for IGBT module degradation data series in [8]



(a)



(b)

Figure 5.3: Plots of baseline clusters for healthy data instances including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only

Table 5.2: Optimal number of clusters and maximum silhouette values of IGBT module degradation data series

Data Source	Max. Silhouette Value	Optimal Number of Clusters
Tounsi (2010)	0.838	4
Scheuermann (2011)	0.927	2
Schmidt (2013) 1E	0.908	2
Schmidt (2013) 2E	0.62	3

Now that the baseline clusters were established, the outlier probabilities associated with each test data instance were computed using SVWD and Fisher criterion (FC), as was discussed in Section 3.3. FC requires establishing two classes and determines their relative separation. In the present work, for each test data instance and each baseline cluster, one class consisted of distances between two healthy data instances within the cluster, while the other consisted of distances between the test data instance and all the healthy data instances in the cluster. The distances in each class were assumed to be normally distributed with the mean and the standard deviation calculated. Then Equation 3.4 was used to obtain the FC value for the cluster and Equation 3.5 was used to compute the probability that the test data instance did not belong to the cluster. Once the probability threshold of 0.995 was reached, the test data instance was considered not belonging to the cluster. If the threshold was reached by all the baseline clusters, the current test data instance was considered an anomaly. Table 5.3 shows the anomaly detection results of all

the IGBT module degradation data series. The anomaly signaling point (ASP) represents the first test data instance that is considered an anomaly. For all the data instances between the onset of test data and the ASP, there was not enough confidence (99.5%) that the test data instance is an outlier to at least one of the baseline clusters. Validation of the technique has been presented in Subsection 3.3.7, so there is strong confidence that the announced anomaly is actually an anomaly.

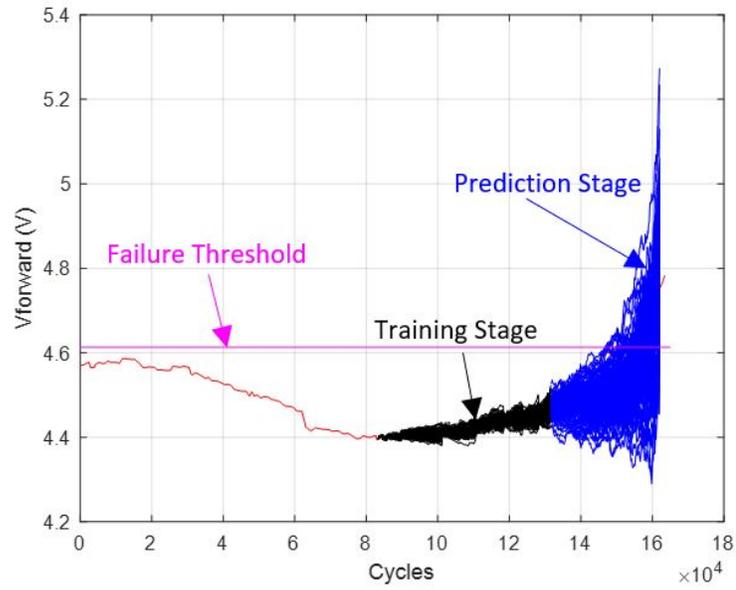
Table 5.3: Points for onset of test data and anomaly signaling point (ASP) of IGBT module degradation data series

Data Source	Onset of test Data (Cycles)	Anomaly Signaling Point (ASP) (Cycles)
Tounsi (2010)	111,200	129,700
Scheuermann (2011)	3,200	5,700
Schmidt (2013) 1E	21,900	56,900
Schmidt (2013) 2E	31,200	53,900

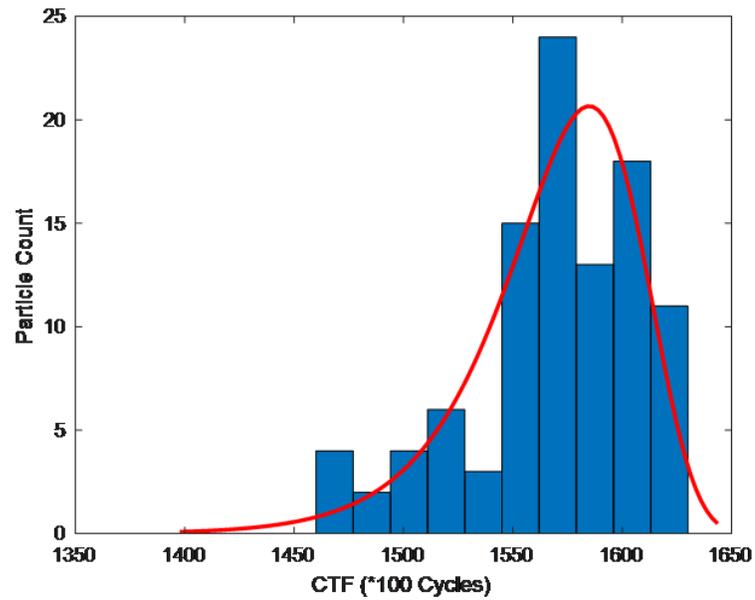
5.3 Results of RUL Estimation Based on SIR Filter

After an anomaly was detected and signaled, the RUL estimation of the device under test (DUT) was triggered. As was discussed in Section 3.3, $l(k)$ was the hypothetic loading variable associated with loading parameters I_{load} and $\Delta T_j(k)$, while $\Delta K(k)$ was the stress intensity factor, as the “driving variable” of the $V_{forward}$ increased. The time-series data of $l(k)$ and $\Delta K(k)$ were derived following Equation 3.22 and 3.14, respectively. Then the SIR filter was implemented for RUL prediction of the devices. The particles were initiated at the 4th sampling point where online

estimation of the system process model parameters was initiated. The noise applied to the process model followed Equation 3.20, in which the magnitude of σ_ω was empirically selected as 0.05. Before an anomaly was detected, the recursive Bayesian procedure including prediction of the particles values at the next time step and update of the particle weights was implemented at each time step, with the resampling step applied to retain the effective particle size. At the anomaly signaling point (ASP), the update of the particle weights was terminated, and the particles were propagated to make unremitting prediction of future states until the failure threshold is reached. Once all the particles reached the failure threshold or the end-of-test point, the estimated RULs presented by the particles were recorded and the distribution of RULs was formed, taking both the RUL values and particle weights into account.



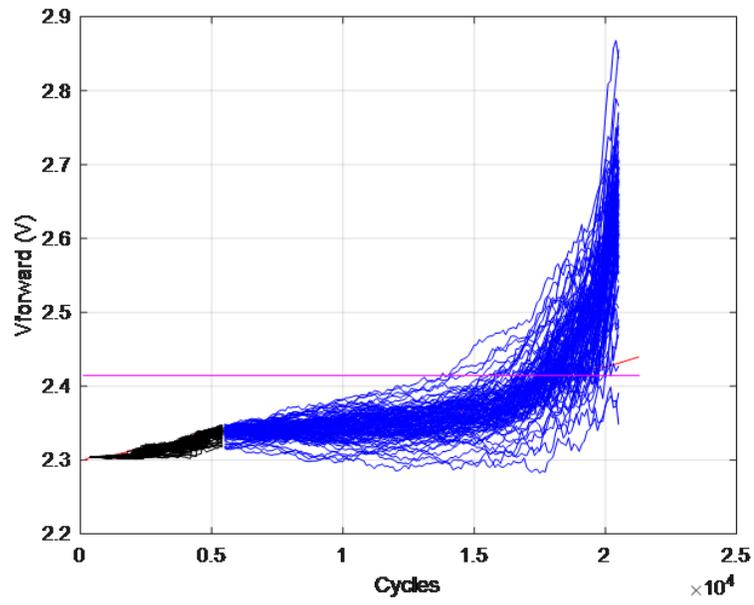
(a)



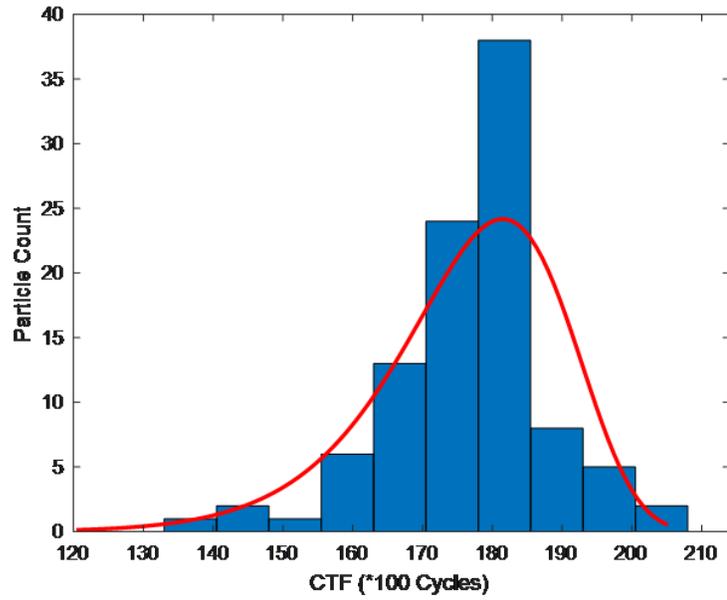
(b)

Figure 5.4: (a) RUL prediction results based on SIR filter; (b) Predicted cycles-to-failure (CTF) distribution of particles for the IGBT module degradation data series in [8]

Figure 5.4 shows the RUL prediction results using the SIR filter for the $V_{forward}$ data in [8], including the histogram showing the distribution of cycles-to-failure (CTF) as a result of variations in particle trajectories. In Figure 5.2 (a), the trajectories of particles before ASP, called the training stage, are marked in black. As an anomaly was detected and signaled at 129,700 cycles, the RUL prediction was triggered, and the particles were propagated until the 5% $V_{forward}$ increase failure threshold was reached, whose trajectories are marked in blue at this stage in the plot, called the prediction stage. The predicted mean CTF calculated from the CTF values and the particle weights was found to be 156,900 cycles, reporting a +3.42% error from the actual CTF of 151,700 cycles, where the raw $V_{forward}$ data reached the failure threshold. To achieve statistical significance of the results given the randomness of noises, the 95% two-sided confidence interval (CI) was also calculated and recorded. Also, to ensure the typicality of the prediction results, 10 individual and independent runs of the prognostic model for each series of degradation data were made to get an average predicted mean CTF and 95% CI. For the degradation data in [8], the average error of the predicted mean CTF was +2.62%.

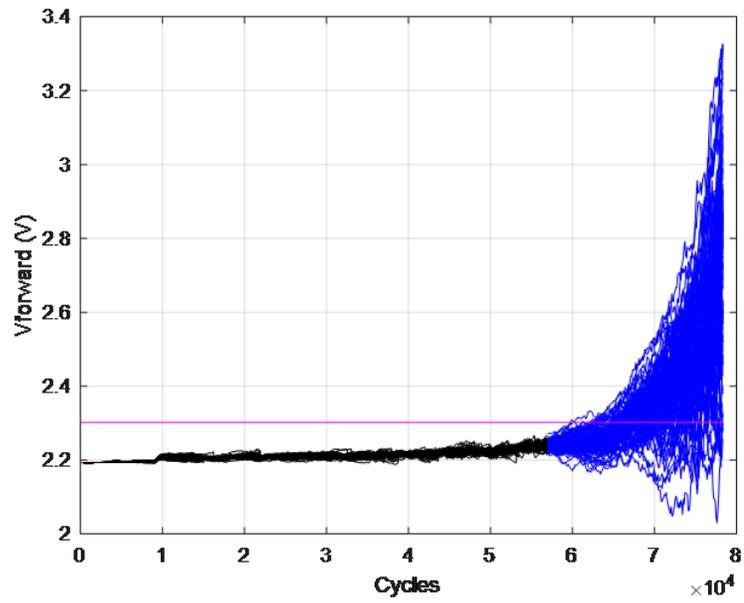


(a)

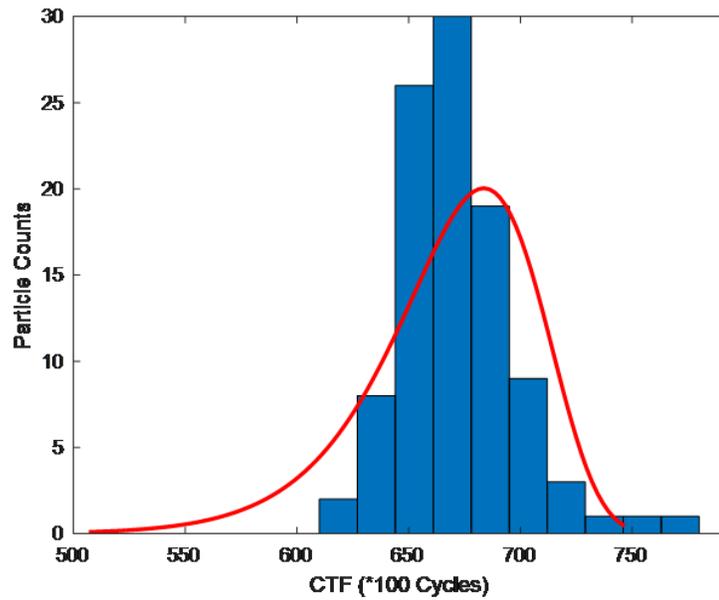


(b)

Figure 5.5: (a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series in [11]

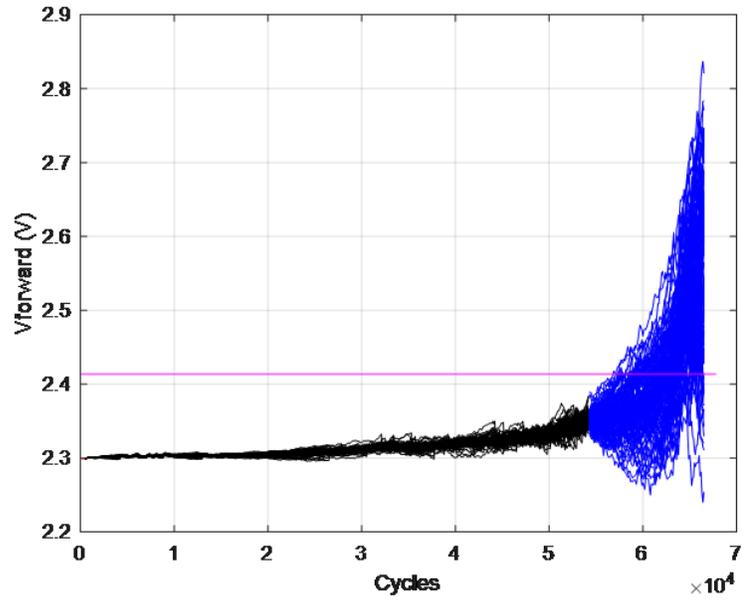


(a)

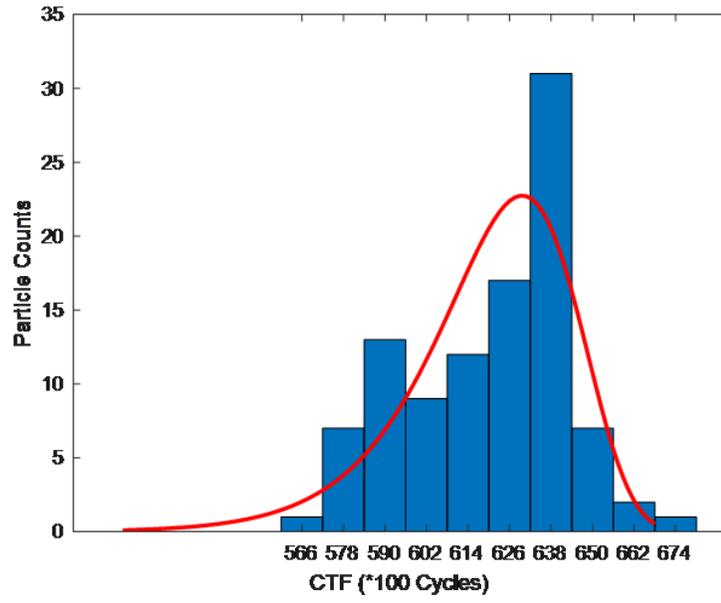


(b)

Figure 5.6: (a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series 1E in [12]



(a)



(b)

Figure 5.7: (a) RUL prediction results based on the SIR filter; (b) Predicted CTF distribution of particles; for the IGBT module degradation data series 2E in [12]

Similarly, the prognostic model was implemented for RUL estimation of the

other IGBT degradation data series, whose results are shown in Figure 5.5-5.7. The average predicted mean errors ranged from -1.71% to +3.53%. Table 5.4 summarizes the results of the average predicted mean CTFs, the raw data CTFs, the average prediction errors and 95% two-sided confidence intervals of all the IGBT module degradation data series studied.

Table 5.4: Summary of RUL prediction results for IGBT modules

Data Source	Average Predicted Mean CTF (Cycles)	Actual CTF (Cycles)	Average Prediction Error (%)	95% Confidence Interval (Cycles)
Tounsi (2010)	157,060	151,700	+3.53	[148,990; 162,110]
Scheuermann (2011)	17,790	18,100	-1.71	[15,830; 19,870]
Schmidt (2013) 1E	66,620	65,500	+1.71	[61,480; 73,370]
Schmidt (2013) 2E	62,350	62,400	-0.08	[57,520; 66,140]

To sum up, the average prediction errors of all the degradation data series were within 4%, which was quite satisfactory, successfully validating the proposed prognostic model for different IGBT modules under various power cycling conditions. A significant advantage of this model is the incorporation of the physics-of-failure knowledge in the system process model which was often empirical in the previous works on PF-based prognostics. The “driving variable” incorporating physics-of-failure knowledge enables more accurate tracking and estimation of the degradation process, combining the merits of both first-principal approaches and real-time monitoring approaches given that a dominant failure mechanism exists, and the physics of this failure mechanism is well understood. Further validation of the prognostic

model will be conducted on GaN HEMTs in the next chapter.

5.4 Resampling Techniques Implementation and Validation

As is discussed in Chapter 4, the conventional resampling techniques, the “classic four”, have severe limitations in that they introduce severe sample impoverishment in the process. Also, resetting the weights to uniformity leads to loss of prior weight information which is undesired. Other resampling methods, including IMHA resampling and roughening resampling, were discussed as well. A novel resampling technique, called IMHA-replacement resampling, was introduced. In this section, this novel resampling technique will be implemented and validated on the same IGBT module degradation data series used for prognostic model validation in the previous chapter. Multinomial resampling, as a representative of the conventional “classic four” techniques, as well as standalone IMHA and roughening resampling, will be used for benchmarking IMHA-replacement resampling with the effective number of particles N_{eff} , the measure of degeneracy, as well as the number of identical particles N_{iden} , which is the difference between the total number of particles and the number of unique particles in the particle set, as a metric of sample impoverishment.

The four resampling techniques in the comparison were fitted in the resampling step of the SIR filter, respectively, with the other parts of the prognostic model being identical. Therefore, variations in the levels of degeneracy and sample impoverishment were purely the effect of resampling. The time-series data of N_{eff} and N_{iden} from the initiation of particles to the anomaly signaling point (ASP) were recorded. Figure

5.8-5.10 show the time-series N_{eff} data from four schemes (no resampling, IMHA resampling, roughening resampling and IMHA-replacement resampling) for the sets of IGBT degradation data in [8], [12]. The comparison based on the degradation data series in [11] is not shown because the number of data instances between particle initiation and ASP is too small to show any evident differences in N_{eff} between the techniques. As can be seen, the N_{eff} values without resampling dropped to a very low level (around 20%) at the ASP, exhibiting a severe level of degeneracy. Standalone IMHA resampling improved the N_{eff} values noticeably, with around 30-40% effectiveness retained at the ASP, but the amount of improvement was far from satisfactory. Both roughening and IMHA-replacement resampling posted a very high effectiveness level of around 90-100%, with the roughening resampling seemingly more perfect (almost 100%). However, this slight edge over N_{eff} does not make a pregnant difference as both are high enough.

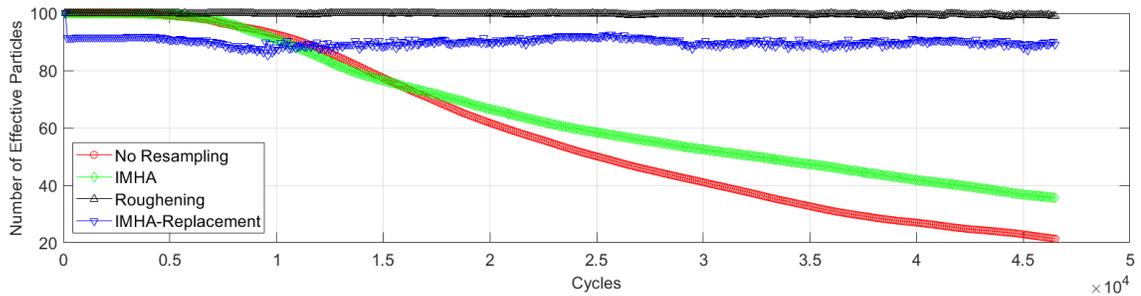


Figure 5.8: Time-series data of N_{eff} for the IGBT module degradation data series in [8]

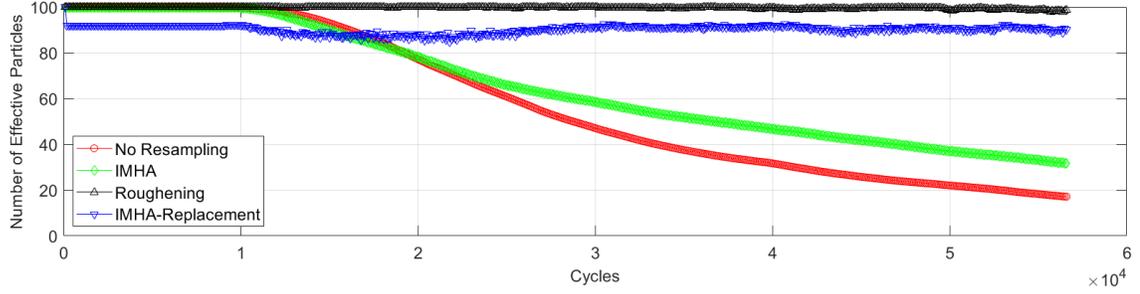


Figure 5.9: Time-series data of N_{eff} for the IGBT module degradation data series 1E in [12]

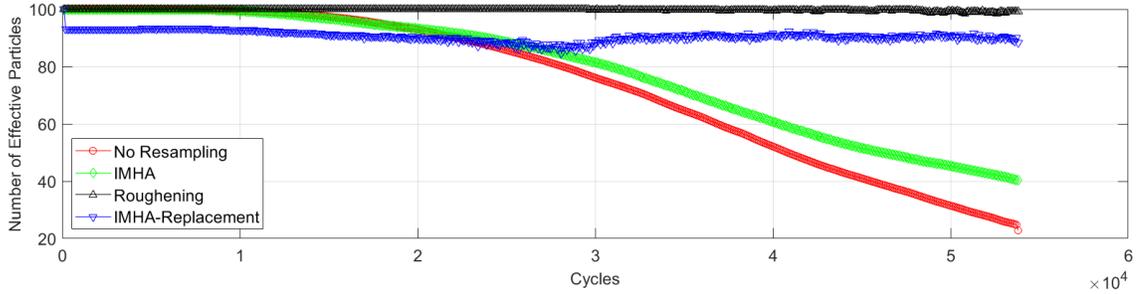


Figure 5.10: Time-series data of N_{eff} for the IGBT module degradation data series 2E in [12]

As for the identical sample size N_{iden} , Figure 5.8-5.10 show the time-series data of N_{iden} for the degradation data using multinomial resampling, IMHA resampling, roughening resampling and IMHA-replacement resampling, respectively. It can be seen that the N_{iden} data for multinomial resampling remained around 40-50% or even higher throughout the process, given that the process noise magnitude ($\sigma=0.05$ V) was quite high compared with the total magnitude of $V_{forward}$ increase (0.11 V to 0.22 V) for the degradation data series. This revealed severe impoverishment of traditional resampling techniques as the number of unique particles was only around half of

all the particles. Standalone IMHA resampling, on the other hand, led to dramatic improvements in the N_{iden} values, with only around 10-20% identical particles most of the time. However, the severe degeneracy of standalone IMHA denied it from being a robust resampling technique. Depending on the cases, roughening resampling offered about 25-50% identical particles in the process which are all lower N_{iden} levels than the corresponding multinomial resampling. Another obvious feature in the N_{iden} data series of roughening resampling was the significant fluctuations of the N_{iden} values. This can be attributed to the additive noise applied in the roughening step, creating greater randomness in the distribution of particles and thus the N_{iden} values. Finally, IMHA-replacement resampling exhibited a much lower (around 10-20% lower), and more stable levels of N_{iden} than roughening resampling. The lower N_{iden} level and the stability of the data gave IMHA-replacement technique a considerable edge over roughening resampling in sample impoverishment. Although the N_{iden} values of IMHA-replacement were slightly higher than that of standalone IMHA, usually less than 5% at ASP, the replacement step was very effective in helping the SIR filter regain high N_{eff} values from 30-40% to 90%. Therefore, it can be concluded that IMHA-replacement is the most robust technique of all the techniques studied for IGBT module degradation data series.

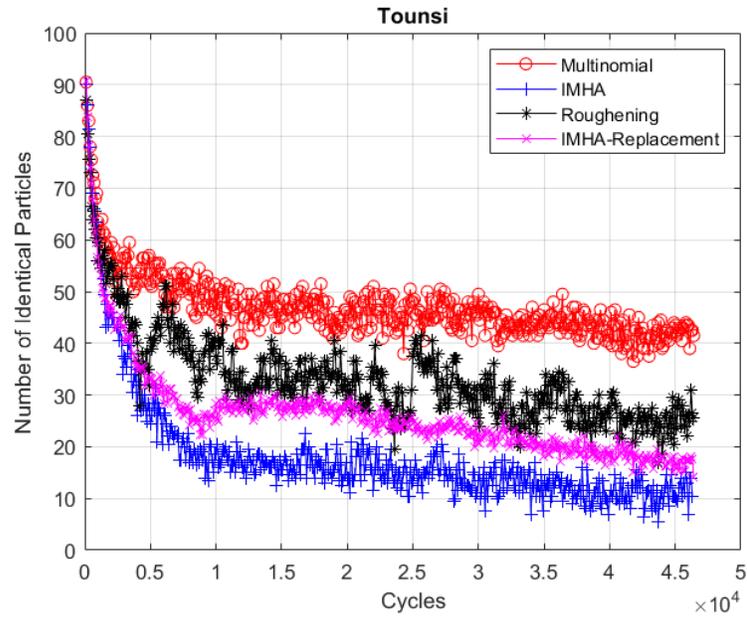


Figure 5.11: Time-series data of N_{iden} for the IGBT module degradation data series in [8]

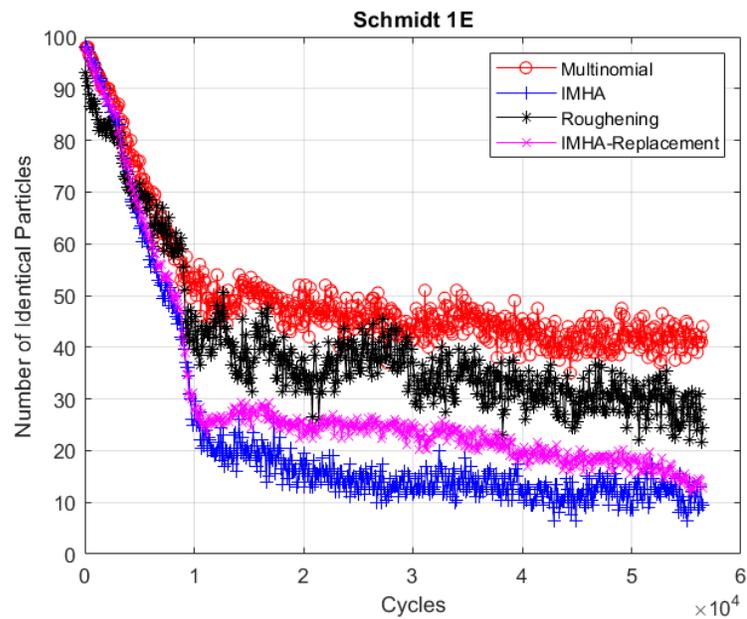


Figure 5.12: Time-series data of N_{iden} for the IGBT module degradation data series 1E in [12]

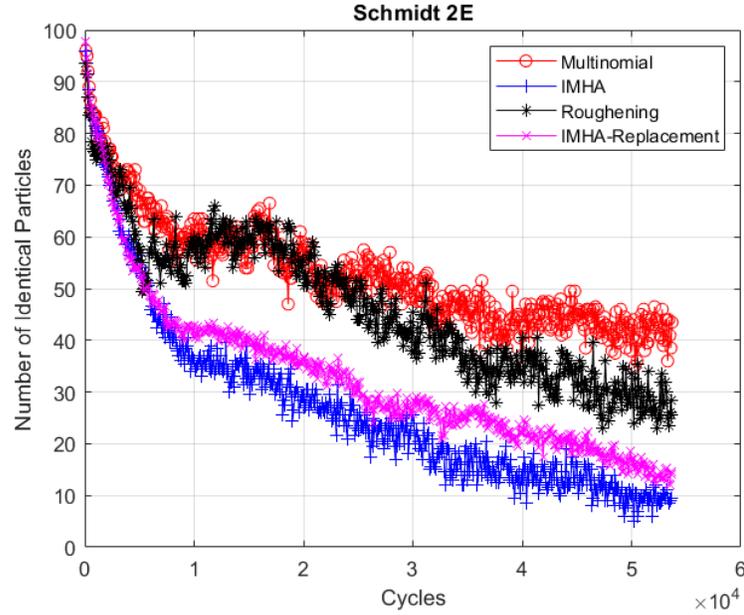
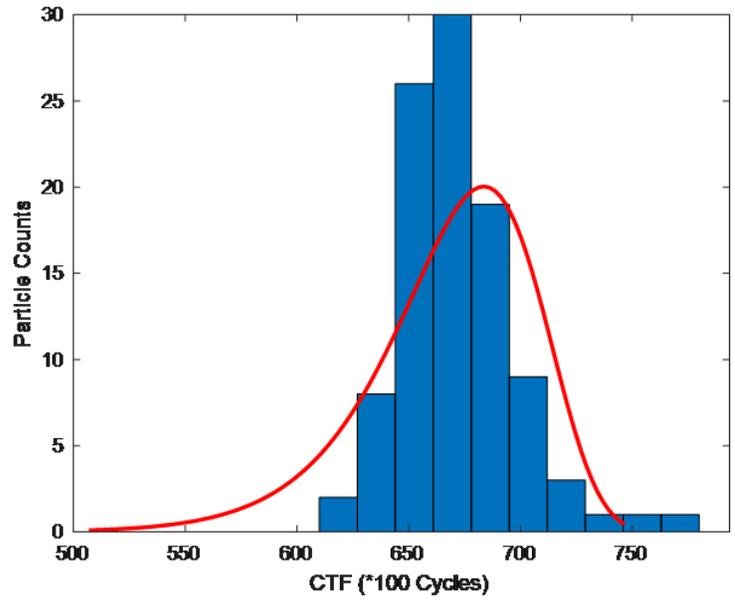
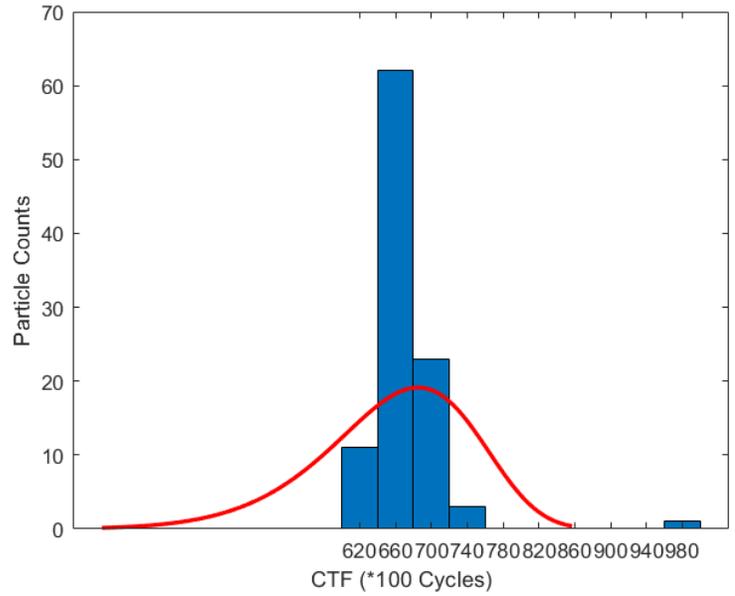


Figure 5.13: Time-series data of N_{iden} for the IGBT module degradation data series 2E in [12]

To reveal the impact of different resampling techniques on the RUL distribution, the RUL distributions obtained from the classic multinomial resampling and the IMHA-replacement resampling for the degradation data series 1E in [12] were compared in Figure 5.14. It can be seen that the ranges and the shapes of the distributions are similar to each other. The IMHA-replacement resampling technique may have produced a tighter and more densely populated particle set, but the impact on RUL distribution seems very limited. Therefore, the scattering of particle trajectories, or the range of RUL distribution, is not correlated to variations in resampling techniques. The determinant factor is the magnitude of the process noise.



(a)



(b)

Figure 5.14: (a) RUL distribution based on the multinomial resampling; (b) RUL distribution based on the IMHA-replacement resampling; for the IGBT module degradation data series 1E in [12]

5.5 Summary

This chapter has presented the validation results of the prognostic model introduced in Chapter 3, as well as the novel IMHA-replacement resampling technique introduced in Chapter 4 on IGBT module degradation data series. The RUL predictions made at the ASP reported an average predicted mean error of below 4%, which was pretty satisfactory. The incorporation of the first-principal model (stress intensity factor and Paris Equation) corresponding to the dominant failure mechanism did provide an accurate estimation of the future evolution trajectories. The proposed IMHA-replacement resampling technique was validated, providing a high level of effective sample size of around 90%, as well as a low sample impoverishment level, with only around 15% identical particles at the ASP. In summary, the robustness of the prognostic model, together with the IMHA-replacement resampling has been completely validated on IGBT module degradation data series.

Chapter 6: Validation of Prognostic Model for GaN HEMTs

6.1 Degradation Data Source

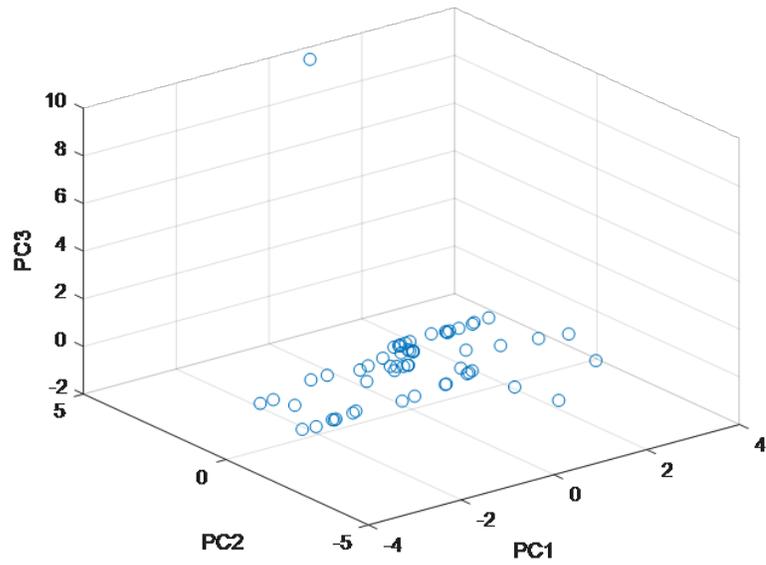
The previous chapter presented the validation of the prognostic model, including the proposed IMHA-resampling technique on IGBT module degradation data series under power cycling conditions. An important contribution of the present dissertation is developing and validating a prognostic model that can be used on different types of power electronic devices with distinct device structures, operating conditions and failure mechanisms. Therefore, this chapter will focus on the validation of the prognostic model and the IMHA-replacement resampling technique on several sets of GaN HEMT time-series degradation data.

Section 1.6 has discussed the degradation mechanism of ohmic contact inter-diffusion of HF GaN HEMTs under HF-ALTs, which was reported in [69]. The tests were conducted on some Wolfspeed G50V3 GaN HEMTs with a nominal drain operating voltage of 50 V. The devices were subject to 200 mA quiescent current and a constant input power of 29 dBm, leading to a nominal output power of approximately 43 dBm and RF gain compression of about 3–5 dB. The junction temperature was maintained constant throughout the test. Several sets of degradation data related to ohmic contact inter-diffusion were presented in this dissertation. Four of them,

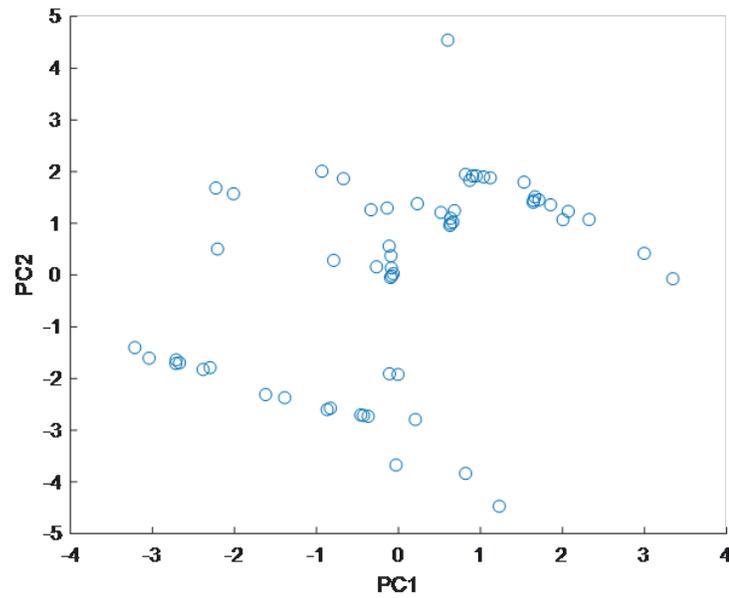
whose corresponding devices were named Device 1 to Device 4, were selected as the raw data for prognostic model validation on HF GaN HEMTs. It should be noted that for these degradation data, the pre-defined failure threshold of -1 dB was never actually reached before the end of test. Therefore, instead of computing the RULs of the devices before the failure threshold was reached, the future state, which was the estimated ΔP_{out} value at the end of test ($t = 800 \text{ hours}$) was predicted and compared with the actual ΔP_{out} value at the end of test.

6.2 Results of Anomaly Detection

The failure precursor of GaN HEMTs is the HF output power degradation ΔP_{out} , while the failure threshold is 1 dB decrease in P_{out} . The “nominal region” was defined using the same method as that of IGBT modules. The same moving window containing 10 latest data instances within the “nominal region” was used to compute the statistical features. PCA was then performed for feature extraction and establishing the feature space. Figure 6.1 shows the 3D and 2D plots of the healthy ΔP_{out} data instances of Device 4 in the feature space based on the first three PCs (PC_1 to PC_3) and the first two PCs (PC_1 and PC_2), respectively.



(a)



(b)

Figure 6.1: Plots of healthy data instances in the PC-based feature space including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only

After feature extraction, k -means clustering was implemented to establish

the baseline clusters for the GaN HEMT healthy data. The silhouette values were observed with k ranging from 2 to 6, just like for IGBT modules. For the selection of k , Figure 6.2 shows the plot of silhouette value variation with k for the degradation data series of Device 4. Table 6.1 summarizes the values of k for the degradation data series of Wolfspeed GaN HEMTs. It can be seen that k still varied from 2 to 4. Figure 6.3 shows the 3D and 2D plots of the healthy data instances from Device 4 partitioned into $k = 4$ clusters, where the points in the same shape and the same color belong to the same cluster. Even though there seemed to be an outlier in the 3D plot with significantly higher PC_3 value, it was still partitioned into a baseline cluster with other points having low PC_3 value. This is because of the relatively low weight of PC_3 in SVWD, leading to the differences in PC_3 being attenuated compared with PC_1 and PC_2 .

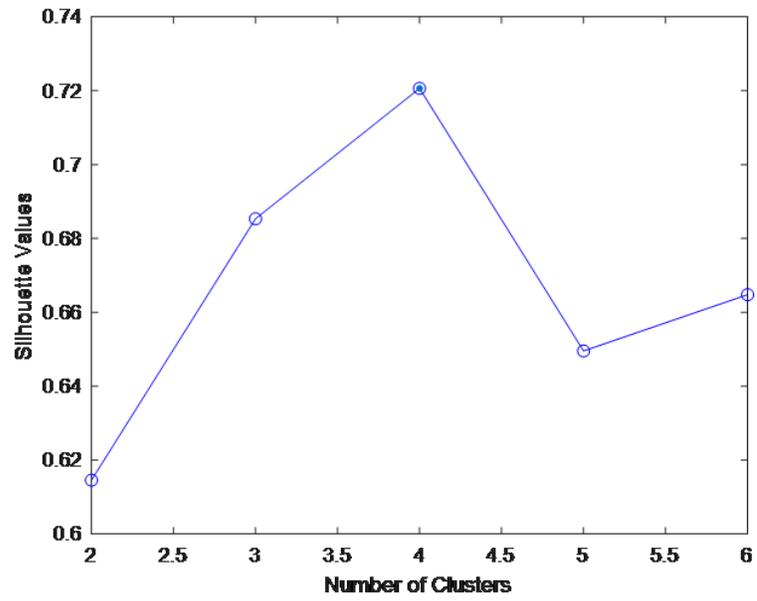
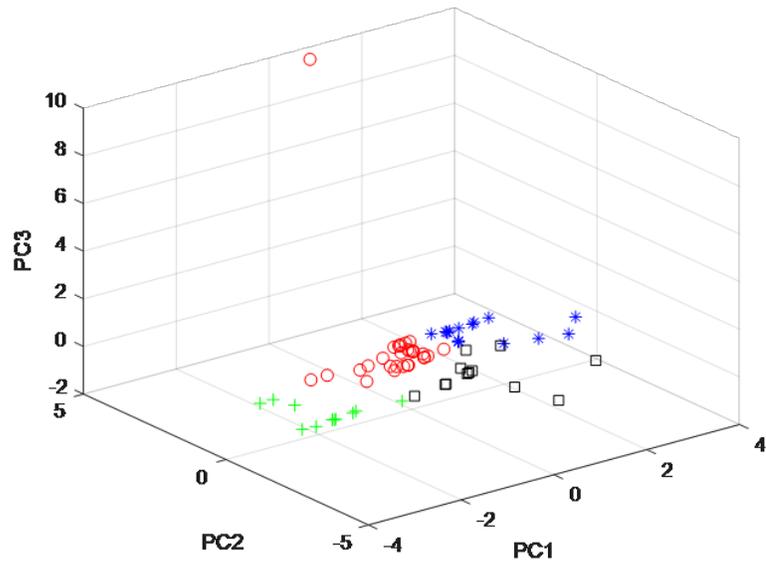
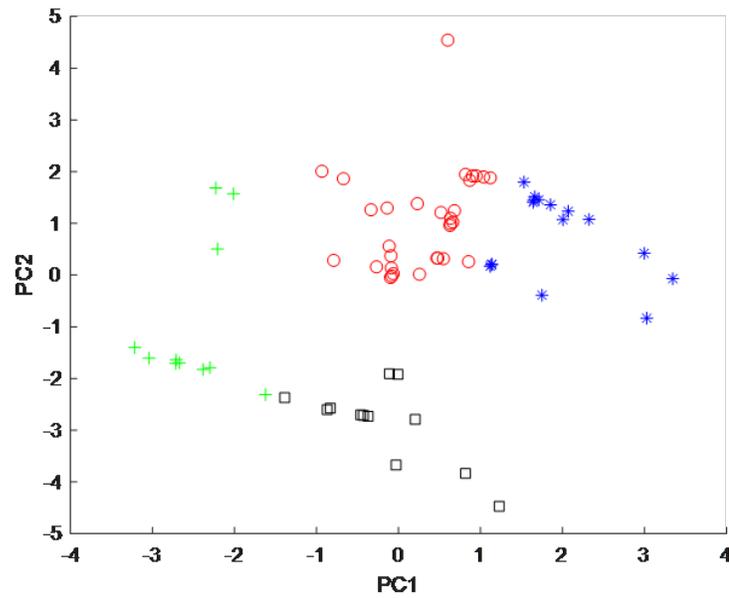


Figure 6.2: Variation of silhouette value with the number of clusters for Device 4 degradation data series



(a)



(b)

Figure 6.3: Plots of baseline clusters for healthy data instances including (a) 3D plot involving PC_1 to PC_3 ; (b) 2D plot involving PC_1 and PC_2 only

As the baseline clusters were established, computation of the outlier probability

Table 6.1: Optimal number of clusters and maximum silhouette values of Wolfspeed

GaN HEMT degradation data series

Device Name	Max. Silhouette Value	Optimal Number of Clusters
Device 1	0.723	2
Device 2	0.912	2
Device 3	0.848	2
Device 4	0.72	4

for each test data instance was performed using Fisher criterion (FC). The two distributions corresponding to each test data instance and each baseline cluster were established. Then the FC value was obtained using Equation 3.4 for the cluster and Equation 3.5 was used to compute the probability that the test data instance does not belong to the cluster. The same criterion for determination of whether a test data instance is an anomaly was used for GaN HEMTs as well. Table 6.2 summarizes the anomaly detection results of all the four Wolfspeed GaN HEMT degradation data series.

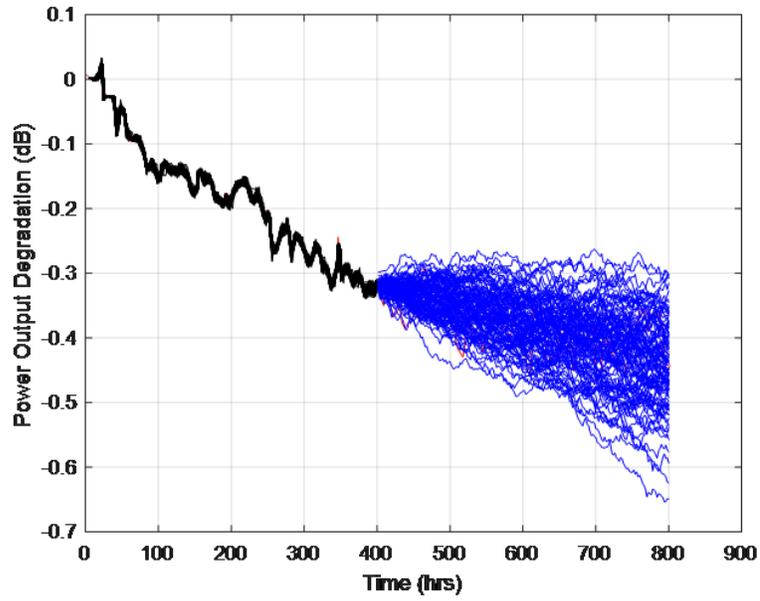
Table 6.2: Optimal number of clusters and maximum silhouette values of Wolfspeed

GaN HEMT degradation data

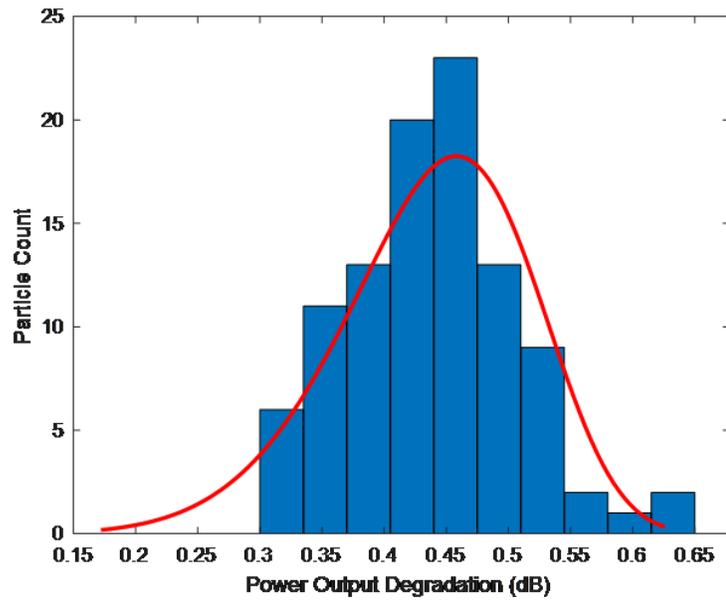
Device Name	Onset of test Data (Hours)	Anomaly Signaling Point (ASP) (Hours)
Device 1	252	400
Device 2	104	128
Device 3	78	110
Device 4	120	300

6.3 Results of RUL Estimation Based on SIR Filter

The RUL estimation of the devices under test (DUTs) can now be performed. For Wolfspeed GaN HEMTs, however, since the pre-defined threshold was not reached before the test ended, predictions of the ΔP_{out} values at the end-of-test point at $t = 800\text{hours}$ were made to compare with the actual ΔP_{out} value. For GaN HEMTs, *load* was the hypothetic loading parameter as a function of HF input power P_{in} and T_j , while the field plate void size $lv(t)$ was the “driving variable” of the ΔP_{out} degradation. *load* was a constant in these cases since P_{in} and T_j were kept constant throughout the tests. The time-series data of $lv(t)$ were obtained from diffusion simulation discussed in the previous section. Then the SIR filter was implemented for prediction of ΔP_{out} at 800 hours. The standard deviation σ_ω of process model noise following Equation 3.20 was empirically selected as 0.01. Otherwise, follow the same procedure of RUL prediction for IGBT modules was followed.



(a)

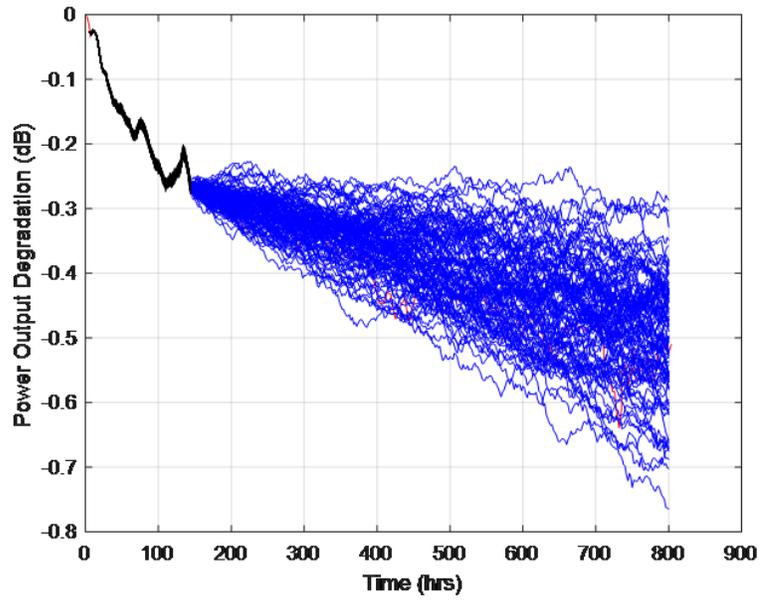


(b)

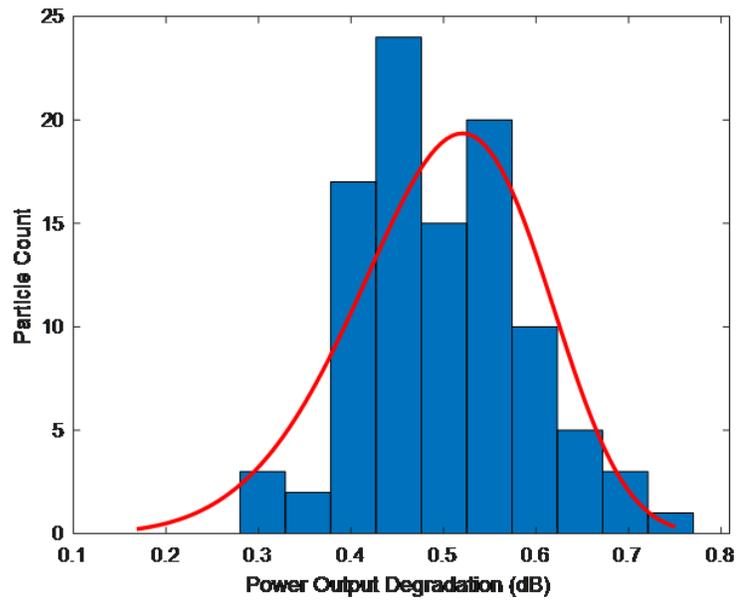
Figure 6.4: (a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at $t=800$ hrs for the degradation data series of Device 1

Figure 6.4 shows the future state prediction results for the degradation data

series of Device 1, including the histogram showing the distribution of predicted states at 800 hours for the particles. As the first anomaly was detected and signaled at 400 hours into the test, prediction of the state at the end of the test was triggered, propagating the particles to $t = 800$ hours. The predicted mean ΔP_{out} based on the values and the weights of all the particles at 800 hours was found to be -0.4292 dB, reporting a +3.88% error from the actual ΔP_{out} value of -0.4465 dB. The 95% two-sided confidence interval (CI) of predicted ΔP_{out} was also calculated and recorded. 10 individual and independent runs of the prognostic model were made for each series of degradation data for an average predicted mean ΔP_{out} and 95% CI. For the degradation data series in [8], the average error of the predicted mean CTF was +3.98%.

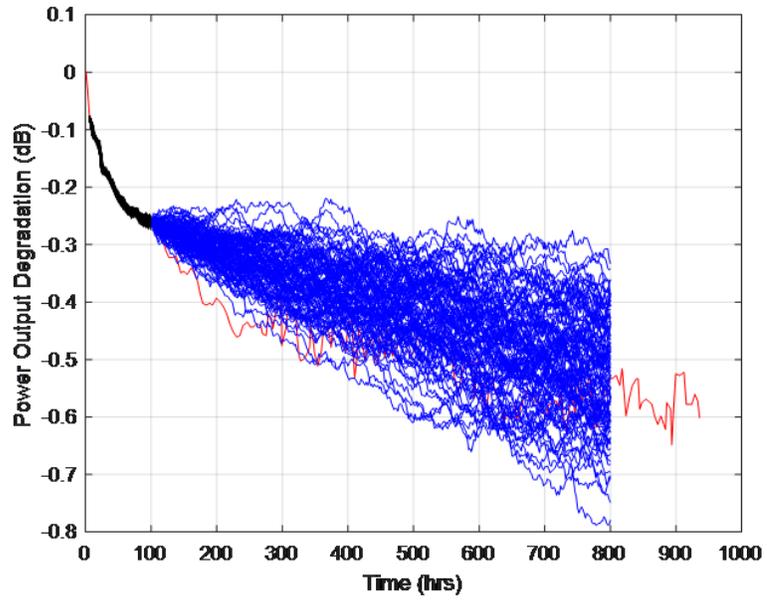


(a)

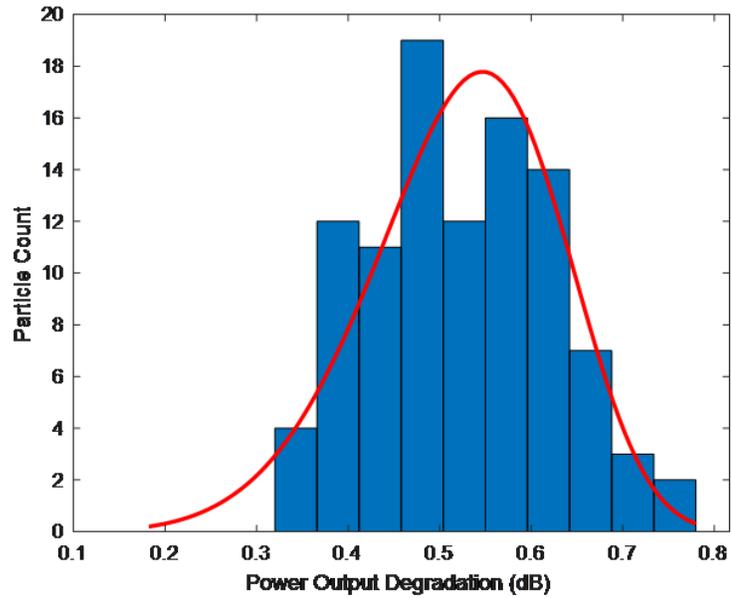


(b)

Figure 6.5: (a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at $t=800$ hrs for the degradation data series of Device 2

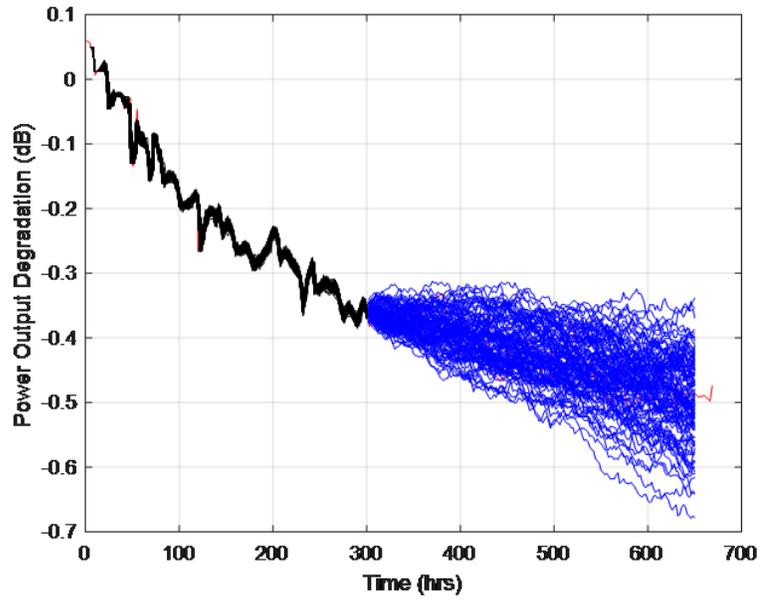


(a)

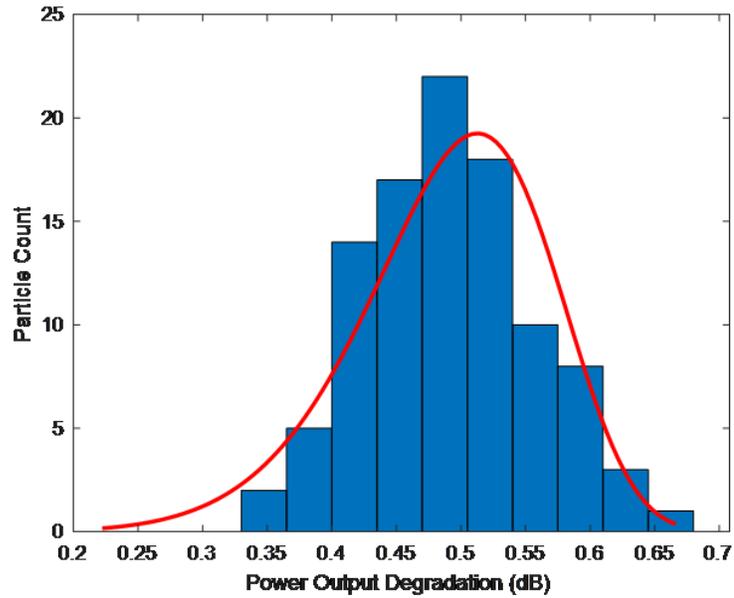


(b)

Figure 6.6: (a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at $t=800$ hrs for the degradation data series of Device 3



(a)



(b)

Figure 6.7: (a) Future state prediction results based on SIR filter; (b) Predicted state distribution of particles at $t=800$ hrs for the degradation data series of Device 4

Figure 6.5-6.7 show the result of end-of-test ΔP_{out} prediction of the other

Wolfspeed GaN HEMT degradation data series. The average predicted mean errors ranged from -0.26% to +6.66%. Table 6.3 summarizes the results the average predicted mean ΔP_{out} , the actual ΔP_{out} in the raw data, the average prediction errors and 95% two-sided confidence intervals of all the four GaN HEMT degradation data series. The average prediction errors of GaN HEMT degradation data series were still satisfactory, all within 7%. So far, the developed prognostic model has been validated on different types of power electronic devices with distinct device structures, under different types of stressors and wear-out failure mechanisms.

Table 6.3: Summary of Future State prediction results for GaN HEMTs

Device Name	Average Predicted Mean ΔP_{out} (dB)	Actual ΔP_{out} (dB)	Average Prediction Error (%)	95% Confidence Interval (dB)
Device 1	-0.4287	-0.4465	+3.98	[-0.5547; -0.2997]
Device 2	-0.4874	-0.5222	+6.66	[-0.6534; -0.3124]
Device 3	-0.5057	-0.5329	+5.10	[-0.7087; -0.2827]
Device 4	-0.4867	-0.4857	-0.26	[-0.6144; -0.3514]

6.4 Resampling Techniques Implementation and Validation

In this section, the IMHA-replacement resampling technique will be validated on the GaN HEMT degradation data series with the effective number of particles N_{eff} , and the number of identical particles N_{iden} .

The resampling techniques compared for IGBT module degradation data series in Section 5.4 were also implemented for GaN HEMT degradation data series,

respectively. The time-series data of N_{eff} and N_{iden} from the initiation of particles to the ASP were recorded. Figure 6.8-6.11 show the time-series N_{eff} data based on four resampling techniques (no resampling, IMHA resampling, roughening resampling and IMHA-replacement resampling) for the degradation data series of Device 1-4. It can be seen that the N_{eff} values without resampling dropped to only around 10-20% at the ASP, which exhibited severe degeneracy. The N_{eff} data of the standalone IMHA resampling dropped even more quickly than the case without resampling, but improved very slightly at the ASP, with around 15-30% effectiveness at the ASP. Both roughening and IMHA-replacement resampling showed a very high level of effectiveness of almost 100%, indicating that both techniques were more than sufficient in maintaining a high level of particle effectiveness.

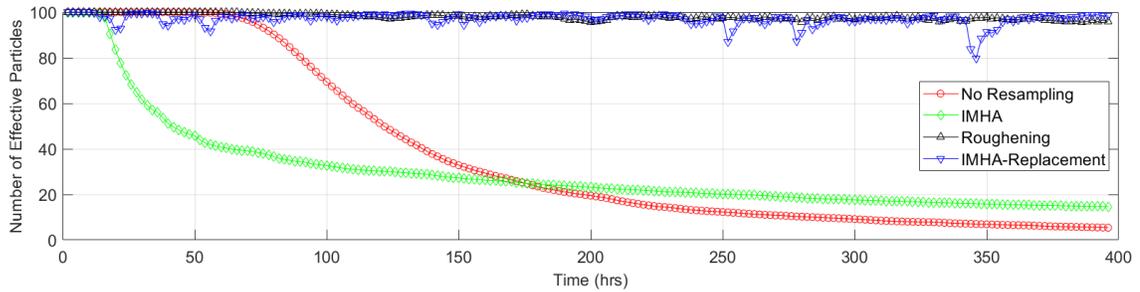


Figure 6.8: Time-series data of N_{eff} for the degradation data series of Device 1

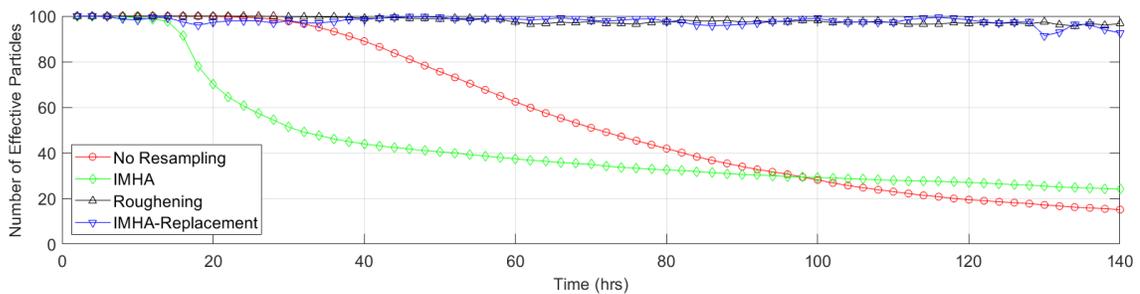


Figure 6.9: Time-series data of N_{eff} for the degradation data series of Device 2

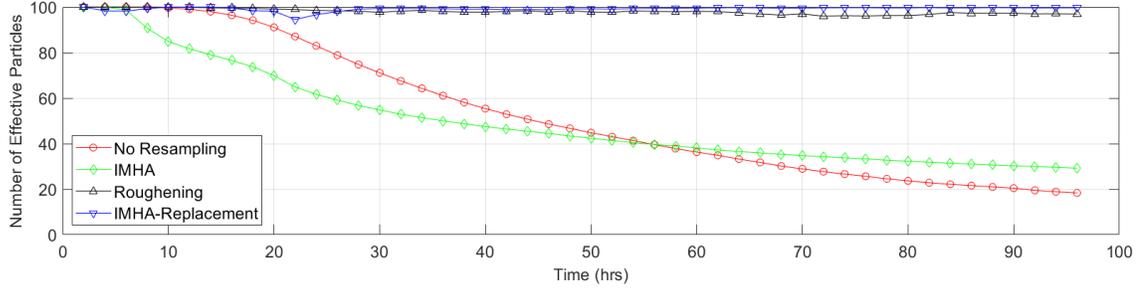


Figure 6.10: Time-series data of N_{eff} for the degradation data series of Device 3

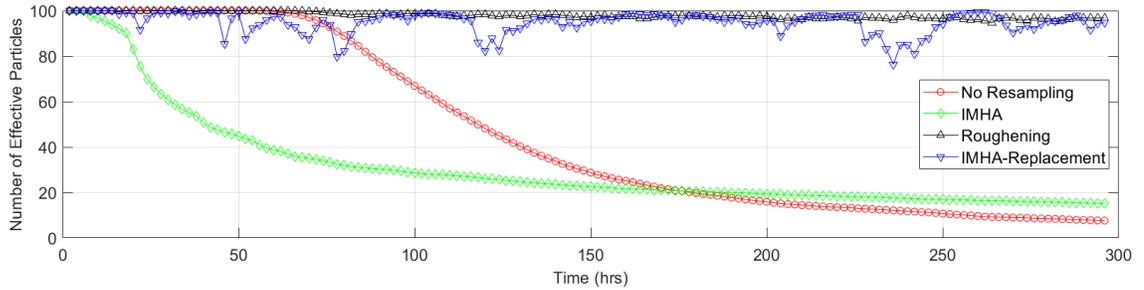


Figure 6.11: Time-series data of N_{eff} for the degradation data series of Device 4

Figure 6.8-6.11 show the time-series identical sample size N_{iden} data of the same degradation data series from the same four techniques as IGBT modules: multinomial, IMHA, roughening and IMHA-replacement. As expected, the N_{iden} data for multinomial resampling remained 40-50% throughout the course, indicating severe impoverishment. Standalone IMHA resampling, like the cases of IGBT modules, also resulted in the lowest N_{iden} values of around 15-25% identical particles most of the time. Meanwhile, roughening resampling posted about 30-40% identical particles in the process lower than multinomial resampling. IMHA-replacement resampling showed about 20-30% identical particles, higher than standalone IMHA but lower than the other two techniques. The relative positions of the N_{iden} values for the four techniques were exactly the same as those of IGBT module data series.

Therefore, the same conclusion in IGBT module data series can be applied, and the robustness of IMHA-replacement resampling was validated on both types of power electronic devices.

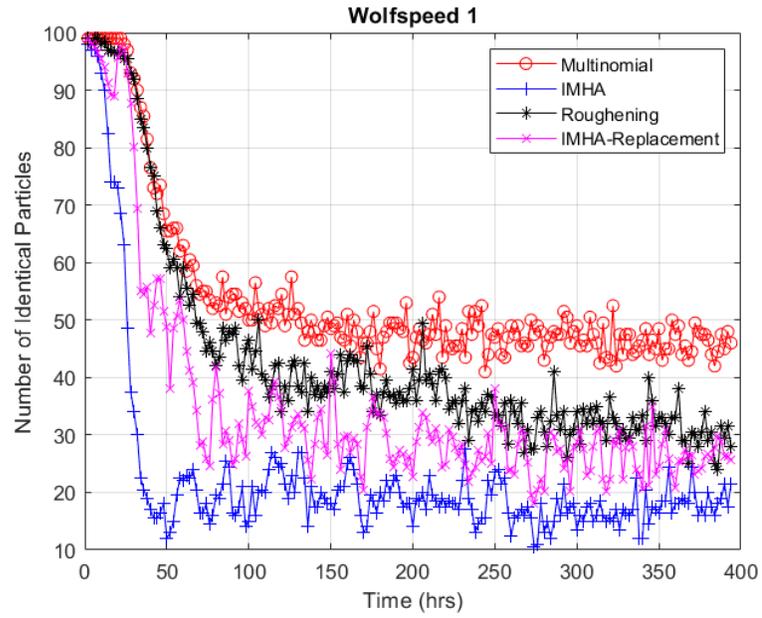


Figure 6.12: Time-series data of N_{iden} for the degradation data series of Device 1

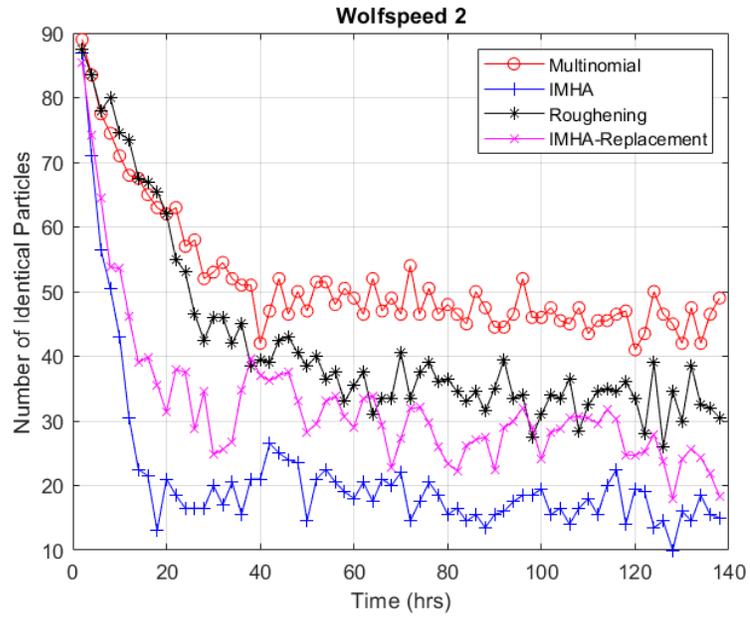


Figure 6.13: Time-series data of N_{iden} for the degradation data series of Device 2

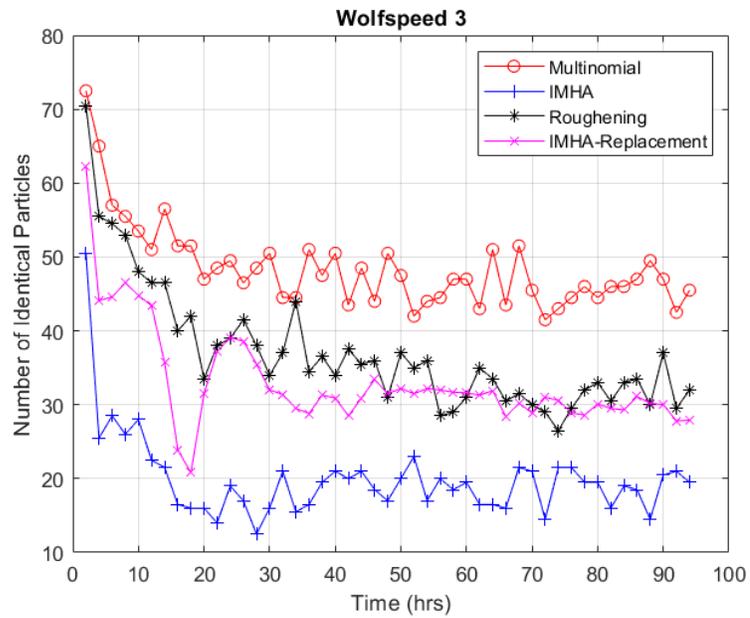


Figure 6.14: Time-series data of N_{iden} for the degradation data series of Device 3

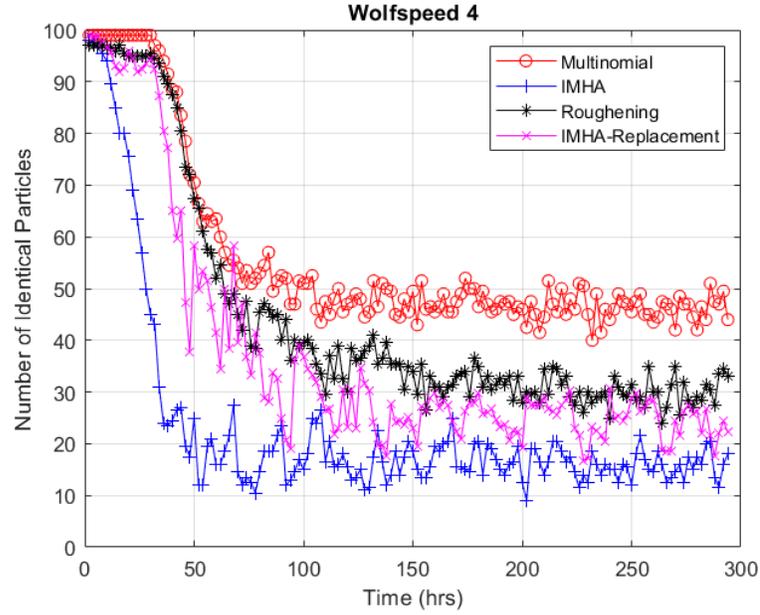
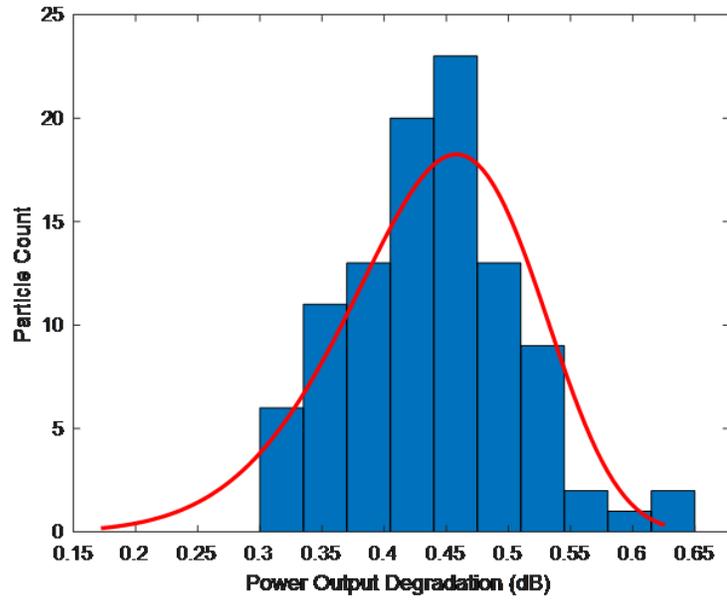
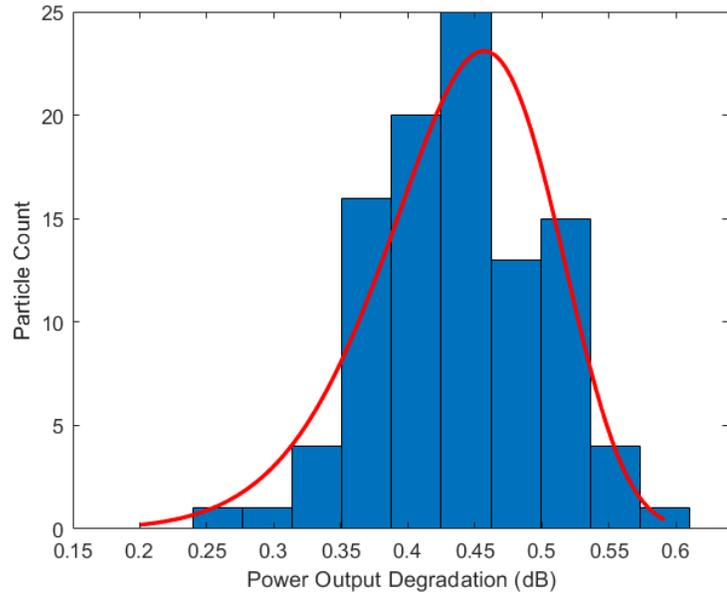


Figure 6.15: Time-series data of N_{iden} for the degradation data series of Device 4

Similarly, the RUL distributions from the classic multinomial resampling and the IMHA-replacement resampling for the degradation data series of Device were compared in Figure 6.16. It turns out that the ranges and the shapes of the distributions are similar as well. The negligible impact of different resampling techniques on RUL distribution was demonstrated again.



(a)



(b)

Figure 6.16: (a) RUL distribution based on the multinomial resampling; (b) RUL distribution based on the IMHA-replacement resampling; for the degradation data series of Device 4

6.5 Summary

This chapter has presented the validation results of the prognostic model and the novel IMHA-replacement resampling technique applied to the Wolfspeed GaN HEMT degradation data series. The dominant failure mechanism was ohmic contact inter-diffusion. The time-series data of void size $lv(t)$ driven by diffusion of Au atoms into the source contact were computed using Mathematica. The future state predictions made at the ASP gave an average predicted mean error of below 7% at $t = 800$ hours, which was also pretty satisfactory. The proposed IMHA-replacement resampling technique was also validated, providing an almost 100% level of effective sample size and a reasonable sample impoverishment level (20-30%). So far, the robustness of the prognostic model and the IMHA-replacement resampling have been demonstrated on IGBT module degradation data series.

Chapter 7: Conclusions and Future Work

7.1 Conclusions and Contributions of The Dissertation

The present dissertation has introduced and validated a novel prognostic model for different types of power electronic devices undergoing a specific type of stressor. Particularly, the present dissertation implemented particle-filter-based prognostics for the first time on GaN HEMT, a representative wide-bandgap semiconductor power device under high-voltage and high-frequency operations. The prognostic model integrated anomaly detection based on semi-supervised machine learning for anomaly detection and remaining useful life (RUL) prediction based on the particle filter (PF) approach. It consisted of two modules: an anomaly detection module and a RUL prediction module including a non-linear process model and a measurement model.

For anomaly detection, semi-supervised machine-learning stood out being a great compromise between the availability of labeled training data and the availability of reference data to enhance prediction accuracy. The establishment of healthy baseline clusters utilized unsupervised machine learning techniques, including PCA for dimensionality reduction and k-means clustering. The “nominal region” of healthy data was determined by measuring the I-V characteristics of a group of

new IGBTs and analyze the fluctuations in their V_{CE} values. The computation of the outlier probability of a test data instance for a baseline cluster was based on Fisher criterion, concerning the distribution of distances between data instances within each baseline cluster, and the distribution of distances between the test data instance and the healthy data instances within each baseline cluster. Finally, this anomaly detection model was validated on the wine dataset, a benchmark dataset for evaluating the performance of clustering algorithms. The new approach outperformed the Mahalanobis-distance-based approach.

The RUL/future state prediction was based on a system process model incorporating a power-law polynomial process model and a measurement model. The key variables of the process model for predicting the future evolution trends were a hypothetical loading variable and a "driving variable". This driving variable depended on the specific type of semiconductor electronic devices, the stressor they were subject to and the resultant dominant failure mechanism. In the present dissertation, the "driving variable" for IGBT modules was the stress intensity factor whose time-series evolution data were derived from Paris equation. For GaN HEMTs, the "driving variable" was the void size in the field plate under high temperature operation of RF-ALT. The time-series evolution data were obtained from solving the interdiffusion equation in Mathematica with the temperature extracted from a temperature profile simulation in COMSOL Multiphysics. The SIR particle filter employing 100 particles were implemented for RUL/future state prediction 10 times for each IGBT module and GaN HEMT degradation data. The prediction results were quite satisfactory, with average errors within 4% for IGBT modules and 7% for GaN HEMTs. The

applicability of the model to different types of power electronic devices, including conventional Si power electronic devices and high voltage high frequency (HVHF) wide-bandgap power devices, was validated.

In terms of particle filter methodology, a novel resampling technique, named IMHA-replacement resampling, was developed and validated in SIR filter. State-of-the-art resampling techniques were reviewed, including the traditional techniques “the classic four”: multinomial, systematic, stratified and residual resampling. The limitations of these resampling techniques were discussed. Taking multiple copies of high-weight particles in these techniques from the original particle set leads to severe sample impoverishment. Besides, increasingly frequent weight-resetting causes undesired loss of prior weight information, and in long-term Bayesian filtering process, this will increase the chance of getting a uniform weight pattern. Some alternative resampling techniques that avoided the limitations in traditional techniques were also reviewed and used for benchmarking, including standalone IMHA resampling and roughening resampling. IMHA-replacement resampling was then introduced, adding a second step after IMHA resampling by replacing the particles having below-average weights with new high-weight particles resampled from the positions of high-weight particles at the previous time step. The IMHA-replacement resampling was validated by outperforming standalone IMHA, roughening and traditional multinomial resampling in the evaluations using the effective sample size and the identical sample number.

Finally, this dissertation also investigated the effect of gamma radiation on the reliability of enhancement-mode (E-mode) GaN HEMTs. A number of commercial

E-mode GaN HEMTs underwent a two-stage gamma irradiation with a wide span of doses from 5 krad (Si) to 60 Mrads (Si). The irradiated devices were characterized using the probe station and semiconductor parametric analyzer to measure shifts in the critical DC characteristics during irradiation, including I-V characteristics, transfer characteristics and gate leakage characteristics. Pronounced degradations were discovered for the transfer characteristics within the first couple of Megarads, while much smaller amounts of degradation were observed for further doses. The gate leakage, on the other hand, almost deteriorated steadily with accumulated total dose. The wide span of irradiation doses helped reveal the complete DC parameter degradation behavior in hard radiation applications of the E-mode GaN HEMTs.

From the above, the scholarly contributions of this dissertation can be summarized as follows:

- Development and validation of a robust prognostic model integrating a novel machine-learning-based anomaly detection technique and particle-filter-based remaining useful life (RUL) prediction.
- Implementation of particle filter for prognostics of high voltage and high frequency GaN HEMTs under HF-ALT, with investigation of radiation hardness under Gamma irradiation (presented in detail in Appendix A).
- Development and validation of the novel IMHA-Replacement resampling technique for the particle filter approach.

7.2 Discussions and Future Work

The prognostic model developed in the present dissertation was validated on two types of power devices with each under one type of stressor (power cycling ALT for IGBT modules, RF-ALT for GaN HEMTs). However, there were a few things defined empirically or with assumptions that may not apply to all the field application circumstances. The first is the process noise magnitude, which was defined empirically in the present work. For field applications, the selection of the noise magnitude depends on the desired RUL distribution. To achieve more conservative RUL predictions, the lower bound of the 95% confidence interval of the distribution has to have a smaller value. i.e. The distribution has to be more scattered, induced by a larger process noise magnitude. If a tighter distribution is required, then a smaller noise magnitude will help. In addition, the failure thresholds of IGBT modules and GaN HEMTs were defined based on empirical assessments. There were, however, no quantitative validations showing that this threshold is the most appropriate. The definition of failure threshold could directly influence the accuracy of RUL prediction and validity of the model. Therefore, in future works, scrutinies of the failure threshold definition based on sensitivity analysis would be very useful. Collection of the degradation data of a large number of relevant devices under the stressor may be necessary to determine if a universal threshold can be identified, and if so, what is the optimal threshold.

Further work can also be done on validating the prognostic model for alternative stressors that the devices may be susceptible to in field applications, such as humidity,

salt spray, vibration or a combination of multiple stressors. The dominant failure mechanism should be identified first, and then an appropriate failure precursor for the failure mechanism needs to be determined in order to implement the prognostic model.

The anomaly detection technique in the present work was based on PCA, k -means clustering and Fisher criterion for two distributions. There are alternative machine learning techniques that could be used for anomaly detection on time-series degradation data, such as Fisher's linear discriminant analysis (LDA). This technique can serve as a tool for dimensionality reduction or a classifier of objects. Benchmarking of the technique with the one developed in the present work, will enable further validation or optimizations of machine learning technique for anomaly detection of time-series data.

In the present work, a single variable was used as the failure precursor. More complex problems may contain multi-variable inputs that require fusion of time-series data from multiple sensors to identify a failure precursor that is not directly measurable. Under multiple stressors, multiple failure precursors may need to be observed individually at the same time as the degradations that lead to different failure mechanisms may develop in parallel. Under these circumstances, the prognostic model may need to be transformed to adapt the specific problems. This is a very meaningful area for further investigation to enhance the applicability of the model.

The radiation effect study of E-mode GaN HEMTs discussed in the appendix can be continued by taking more measuring points at lower doses, so that a larger number of data instances will be available for implementation of the prognostic model.

Future works on radiation effect of semiconductor power devices should focus on low-dose radiation (a few rads to several krads) to stick closer to the actual radiation conditions in field applications. In this sense, the PF-based prognostic model in the present work can be implemented to predict when the parametric degradation will reach the maximum allowed degradation level. Furthermore, the degradation mechanism has to be well understood in order to select a proper “driving variable” for the model.

The prognostic modeling introduced in the present dissertation focused on power electronic devices. However, the application domain of this prognostic model is not confined in power devices. Other electronic components such as electrolytic capacitors, mechanical components such as the rotor blades of wind turbines, as well as other applications such as real-time monitoring of the quality of lubricants may also use this prognostic model for RUL prediction. In fact, the failure of IGBT modules package studied in the present work was mechanical failure (cracking) induced by fatigue. As long as the dominant failure mechanism can be identified and belongs to wear-out failure mechanisms, the model should be able to provide a robust lifetime prediction. In one word, a broader application domain of the prognostic model can be anticipated in the future work.

Appendix A: Gamma Radiation on Enhancement-mode GaN HEMTs

A.1 Introduction

A.1.1 Gamma Radiation

Gamma radiation is a type of penetrating electromagnetic radiation. It is one of the three types of radioactivity, with the other two being alpha and beta radiation. Gamma rays have the smallest wavelengths ($< 0.1 \text{ nm}$) and the highest energy in all waves in the electromagnetic spectrum [149]. On earth, gamma rays are generated from the hottest and most energetic matters or events, such as nuclear explosions, lightning, and the less spectacular activity of radioactive decay of atomic nuclei [150]. An atom is considered unstable if the binding energy is not strong enough to bond its nucleus together, causing imbalanced energy inside the nucleus. As a result, the unstable nucleus spontaneously disintegrates, emitting electromagnetic rays. This process is called radioactive decay.

^{60}Co is a common example of gamma radiation, which first decays to excited ^{60}Ni through emission of an electron. Then the excited ^{60}Ni decays further to the ground state by emitting gamma rays [151]. Another example is ^{241}Am , which mainly decays to ^{237}Np emitting alpha rays. However, gamma rays are emitted as a

byproduct as well. ^{192}Ir is also an example, decaying by emitting beta and gamma rays producing ^{192}Pt [152].

A.1.2 Enhancement-mode Devices

Enhancement mode (E-mode) GaN HEMTs are devices with a positive threshold voltage ($V_{th} > 0$), or normally-off characteristic, which is realized by raising the conduction band level underneath the gate. E-mode devices are valuable for RF and microwave applications because they allow elimination of negative-polarity power supply, simplifying the circuit design [153]. In power electronics, E-mode devices facilitate safe mode of operation which is highly desirable. Other advantages include higher energy and power efficiency, and higher reliability [154]. There are many technologies proposed to realize E-mode operation of GaN HEMTs, including recessed gate structures [155], [156], [157], fluorine treatment [158], [159], high k dielectrics and multi-cap layers [160], p-type layer underneath gate [161], [162], piezoneutralization layer [163] and nonpolar a-plane channel [164]. The fundamental difficulties in the realization of E-mode operation are the technical issues of wet etching in GaN due to the material's strong chemical inertness, as well as the strong polarization effects in AlGaN/GaN heterostructure causing increasing negativity in threshold voltages [153]. The most common technology, recessed gate structure, suffers from mobility and transconductance deterioration of the devices due to the severe dispersion effect in etch-exposed channel [156], [165].

A.1.3 Radiation-induced Degradation of GaN HEMTs

GaN HEMTs have shown great promises in space, defense and nuclear applications, which often involve radiation-abundant environments that require high radiation tolerance to fluxes of neutrons or gamma ray of devices. To turn the promising potential of GaN HEMTs into reality, investigations on the influence of irradiation on the device performance and reliability are essential. Many studies have investigated the radiation hardness of group III–nitride compound semiconductor devices under various irradiation types, such as electron [166], proton [167], [168], [169], neutron [170], and X-ray [171], and various total irradiation doses. They generally showed higher radiation tolerance compared with Si because of high displacement energy of GaN [172], [173], demonstrating the favorability of GaN devices in hard radiation environments.

Many studies have investigated the influence of ^{60}Co gamma radiation on device electrical characteristics. Aktas et al. [174] examined the variations in the characteristics of SiN passivated GaN HEMTs under ^{60}Co -induced gamma radiation up to 600 Mrads. Limited responses to the high radiation doses, including monotonic variations in the DC characteristics with increasing doses, as well as shifts in threshold voltage and maximum transconductance were observed. There were no significant changes in the high-frequency characteristics, the sheet carrier density, the mobility or the contact/sheet resistance of the samples. Belyaev et al. [175] irradiated lab-fabricated GaN HEMTs under a total dose of 10^9 rads. It was discovered that noticeable changes in DC characteristics occurred at lower total doses and depended

on HEMT topology. All the parameter variations at the highest dose were within 20%. Kim et al. [176] irradiated InAlN HEMTs with up to 500 Mrads doses of gamma rays. Significant degradations in electrical properties were discovered after the radiation, and the magnitudes were much larger compared with AlGaN HEMTs. Anderson et al. [177] attempted to induce defects through electrical and radiation stress with gamma radiation of up to 2 Mrads. However, less than 5% degradation in drain-source saturation current I_{DS} was found. Smith et al. [178] reported the results of putting InAlN and AlGaN HEMTs under 9.1 Mrads of ^{60}Co gamma radiation. For devices without surface capping, InAlN HEMTs exhibited greater stability under off-state bias stressing and gamma irradiation in DC characteristics. However, InAlN HEMTs capped with plasma-enhanced chemical vapor deposition (PECVD) SiN_x surface passivation were more susceptible to trap-related degradations in radiation hardness compared with AlGaN HEMTs with the same passivation layer, because of increasing carrier captured in traps at the InAlN/ SiN_x interface.

A.2 Devices Under Test (DUTs)

The irradiated devices in the present work were EPC2035 eGaN FETs from EPC Corporation. The devices were supplied in the form of 0.9 mm square passivated dies with round solder bumps for contacts. They have a rated continuous drain-to-source voltage V_{DS} of 60 V, and a rated continuous drain current I_D of 1 A [179]. The devices were designed for applications including high speed DC-DC conversion, wireless power transfer, LiDAR, etc. These GaN-on-Si devices were fabricated on Si

wafers considering process compatibility and cost. An aluminum nitride (AlN) thin layer was grown on the Si wafer to be a seed layer for the AlGaN/GaN heterostructure. Subsequently, the AlGaN/GaN heterostructure was grown on the AlN layer, with a thin AlGaN layer grown above the highly resistive GaN serving as a strained interface between the GaN and AlGaN crystal layers, allowing the creation of a two-dimensional electron gas (2DEG) filled with abundant free electrons with ultra-high mobility [180]. The gate electrode was processed subsequently forming a depletion region under the gate. As an E-mode transistor, the device is turned on by applying a positive bias to the gate. To lead the electrons to the gate, drain and source contacts, multiple layers of metal were placed through the insulating layer with passage interconnections.

A.3 Condition and Procedure of Gamma Radiation

All the irradiations were conducted at room temperature by ^{60}Co gamma-rays with a flux of 318.5 Rads (Si)/s, at the dry-cell panoramic gamma irradiator in the University of Maryland Radiation Facilities. The entire irradiation procedure consisted of two parts. The first part was higher-dose irradiation with total doses of up to 60 Mrads based on two stages. At the first stage, 9 devices were irradiated for 10 Mrads, while at the second stage, the devices were irradiated for another 50 Mrads, with three of them taken out after 25 Mrads during the second stage. Therefore, three out of the nine devices, #1, #5 and #6, received a total dose of 35 Mrads, and the other six, #2 – #4 and #7 – #9, got 60 Mrads. The second part was lower-dose

irradiations with total doses of up to 2 Mrads. Two groups of devices, each of which had 3, were subject to different doses, with one group, #10 – #12, irradiated for 600 krads, and the other group, #13 – #15, irradiated sequentially for cumulative doses of 5 krads, 20 krads, 100 krads and 2 Mrads, respectively. The temperature throughout the irradiation process was maintained below 30 °C. All contacts were grounded during the irradiation.

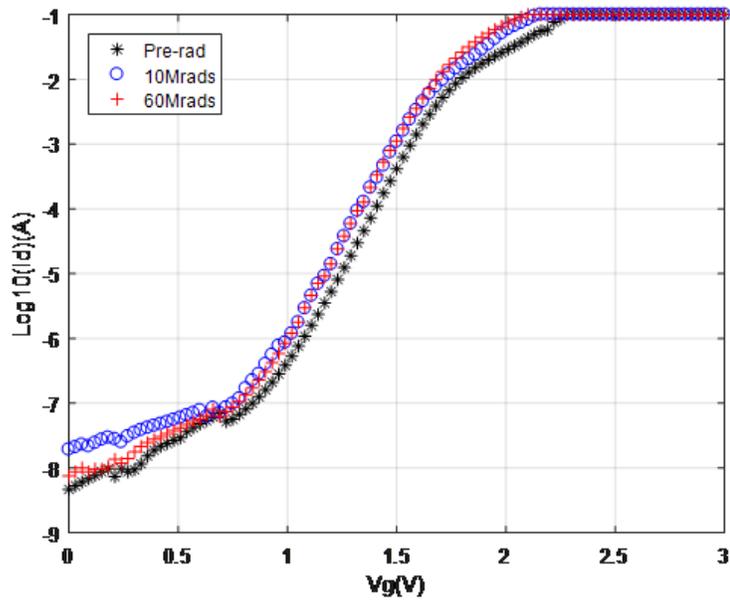
A.4 Electrical Characterization

The DC characteristics of the tested devices were measured at a Micromanipulator probe station with adjustable needle probes, coupled with an HP 4155C semiconductor parameter analyzer. Three critical DC parameters were extracted from the measurements, including threshold voltage V_{th} , maximum transconductance $g_{m,max}$, and gate leakage current $I_{g,leak}$. To minimize the errors induced by contact between needle probes and device solder bumps, multiple independent measurements were taken for each parameter to clear off the possibility of outliers.

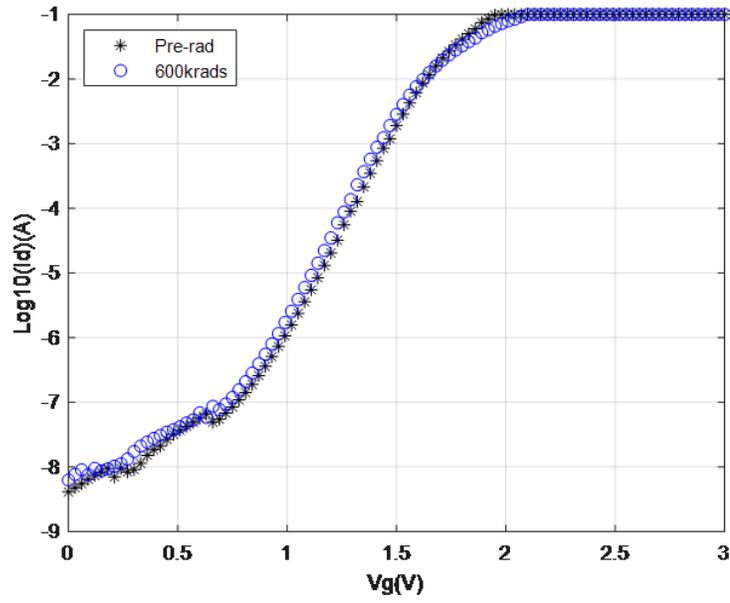
A.5 Results of Device Characteristics Pre- and Post-Irradiation

The DC characteristics measured pre- and post-irradiation were plotted for comparison. The shifts of transfer characteristic curves for device #7 and #13, which were irradiated for 60 Mrads and 600 krads, respectively, are shown in [Figure A.1](#). For the high-dose-irradiated device #7, there was a conspicuous negative shift after the first 10 Mrads of radiation. Very little degradation, if any, occurred at

the second stage. In the case of lower-dose device #13 (600 krad), the negative shift in transfer characteristics was slighter but still noticeable. Figure A.2 shows the leakage current variations of device #5 and #13 with a total dose of 35 Mrads and 600 krad, respectively. Conspicuous increases in $I_{g,leak}$ were observed for both stages in high-dose irradiation, showing an increasingly severe gate leakage as the cumulative dose increased. On the other hand, low dose irradiation of 600 krad only led to a bit increase in $I_{g,leak}$.

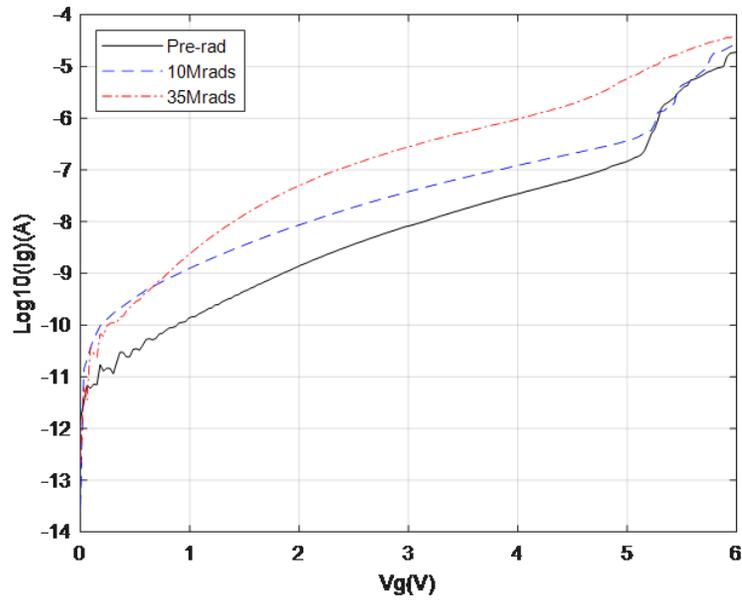


(a)

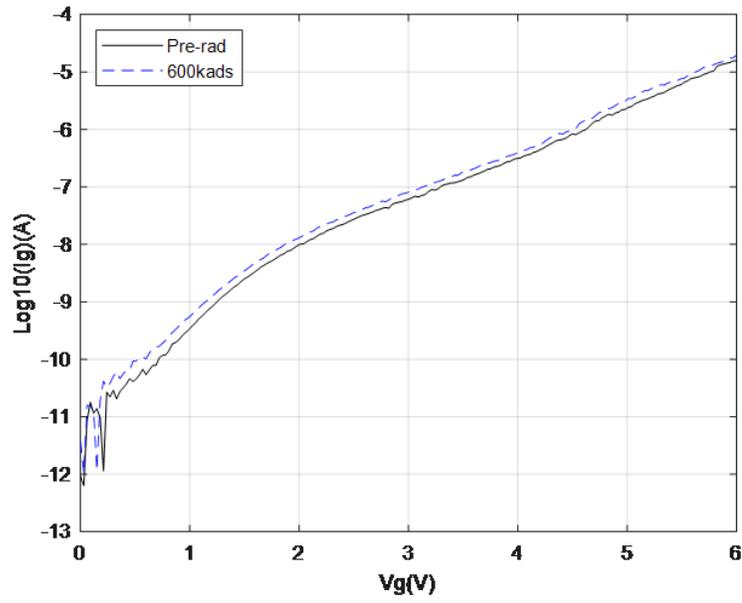


(b)

Figure A.1: Transfer Characteristics of (a) Device #7 and (b) Device #13 Pre- and Post-irradiation



(a)



(b)

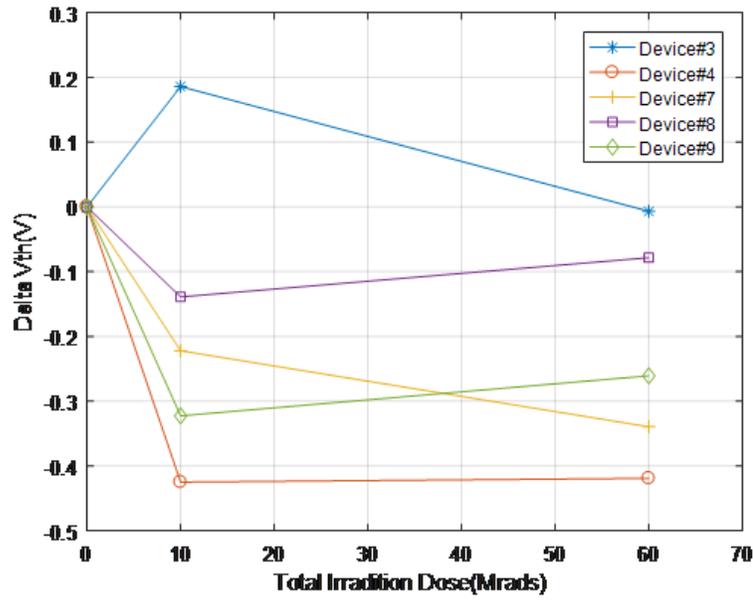
Figure A.2: Gate Leakage Current of (a) Device #5 and (b) Device #13 Pre- and Post-irradiation

The values of V_{th} , $g_{m,max}$ and $I_{g,leak}$ were then extracted from the raw data

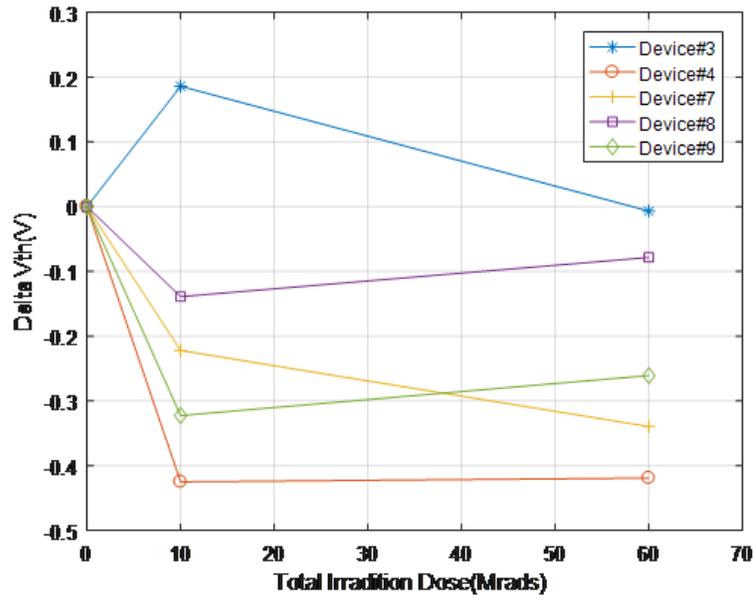
for comparison. The values of $I_{g,leak}$ were extracted at $V_g = 3V$. For V_{th} , the pre- and post-irradiation comparisons were made based on the absolute magnitude of shift ΔV_{th} from the pre-irradiation values. Meanwhile, the post-irradiation values of $g_{m,max}$ and $I_{g,leak}$ were normalized to the pre-irradiation values to gauge the amount of variations. Parametric variations within each group of devices receiving the same total dose were compared, for all the device groups in the present work. Figure A.3-Figure A.8 show the device-to-device variations of V_{th} , $g_{m,max}$ and $I_{g,leak}$, respectively for the group of devices receiving 60 Mrads in total (#3, #4, #7 – #9) and the group receiving sequential doses of up to 2 Mrads (#13 – #15), with the results only shown up to 100 krad for the illustration purpose. For the higher-dose group, the results of #2 are not shown because the device failed to exhibit valid transfer characteristics after the irradiation and was considered to have failed during the irradiation. From the plots, it is evident that the trends of variation agreed well between devices, but the device-to-device variations were quite large. The principal reason was believed to be the variation of the device DC characteristics themselves. Other reasons may include measurement errors incurred mainly due to the instability of contact resistance between the needle probes and the solder bump contacts on devices.

First, the degradation behavior of the 60 Mrads group was observed. For V_{th} , in agreement with the observations in Figure A.1, most of the devices exhibited a relatively significant decrease ranging from 0.1-0.5 V at the first stage of 10 Mrads, and then at the second stage it decreased much less or even increased slightly. If excluding device #3, which showed an anomalous increase in V_{th} after 10 Mrads, the

average decrease after 10 Mrads was 0.2771 V. After 60 Mrads, the average decrease was 0.2744 V, which was even a bit less than that after 10 Mrads. For $g_{m,max}$, a similar pattern was shown with relatively significant average decrease of 36.49% observed in the first 10 Mrads and only slight decrease or even increase (averaged 6.14% further decrease) in the next 50 Mrads. For $I_{g,leak}$, nevertheless, all the devices exhibited persistent increases throughout the radiation with total magnitude between 10 and 40 times (averaged 27.66 times increase), while the increases occurred at the first stage averaged 6.88 times, accounting for 24.89% of total increase after 60 Mrads. Generally, the degradations of the DC parameters were more significant in the beginning part of high-dose irradiations.



(a)

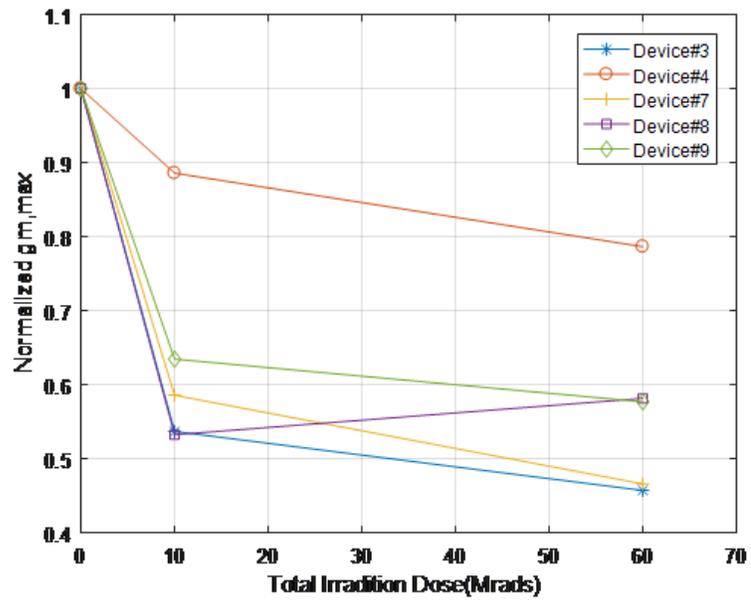


(b)

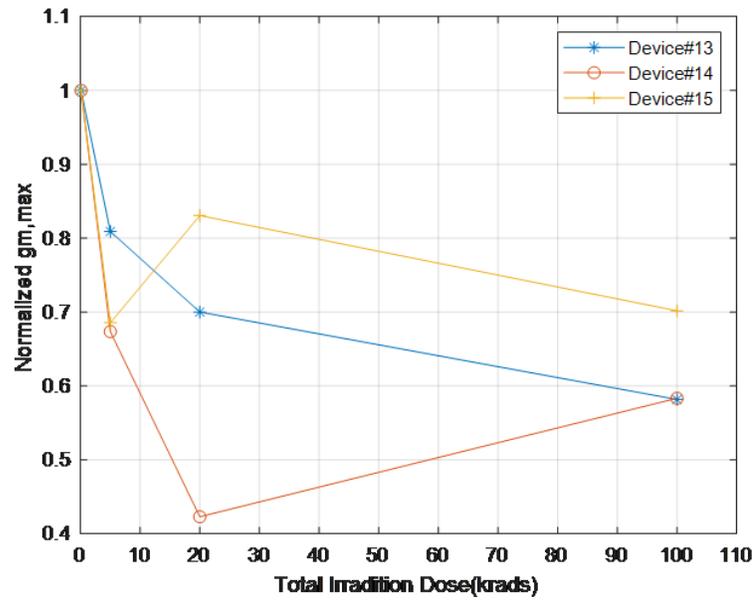
Figure A.3: Device-to-Device Variations of Threshold Voltage in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group

As for the sequential 2 Mrads group, similar variation behavior was observed.

For V_{th} , the average decreases at 5 krad, 20 krad, 100 krad and 2 Mrad were 0.0824 V, 0.128 V, 0.115 V and 0.1655 V, respectively. The increase from 20 krad to 100 krad was due to the increase from device #18. For $g_{m,max}$, the average degradations at 5 krad, 20 krad, 100 krad and 2 Mrad were 27.73%, 34.88%, 37.78% and 46.09%, respectively. The average degradation at 2 Mrad was even higher than after 10 Mrad in the 60 Mrad group, which was understandable considering the large variability between devices. For $I_{g,leak}$, the average normalized increases at 5 krad, 20 krad, 100 krad and 2 Mrad were 53.94%, 69.92%, 86.8% and 115.44%, respectively. Compared with the 60 Mrad group results, the increase in $I_{g,leak}$ was more consistent from low to high doses. Overall, the average magnitudes degradation of all the three parameters increased steadily in the 2 Mrad total dose, with surprisingly considerable degradation observed for the first 5 krad.

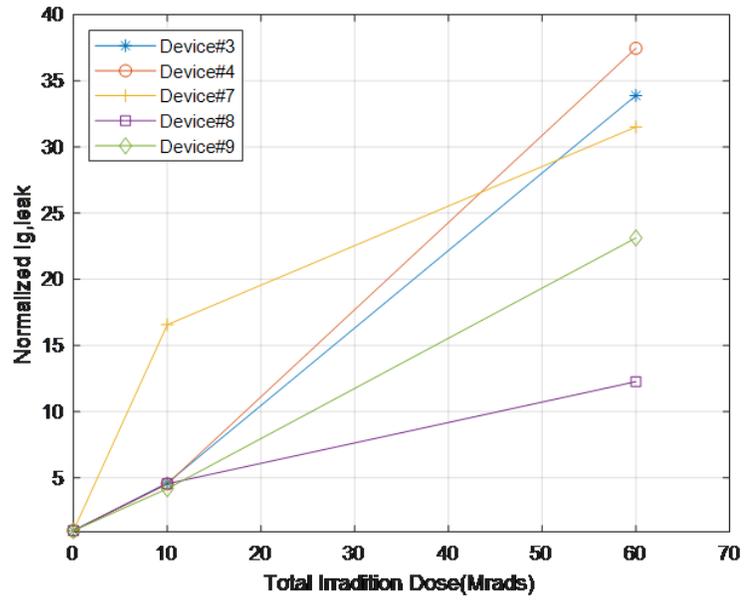


(a)

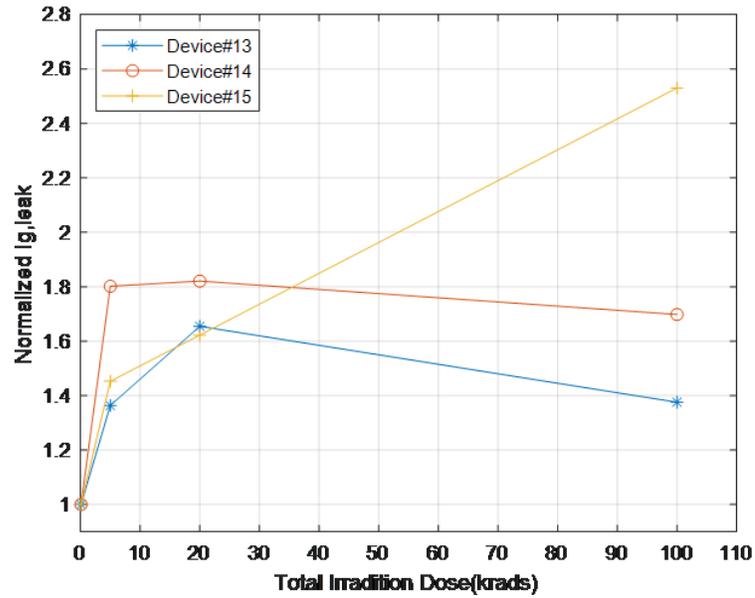


(b)

Figure A.4: Device-to-Device Variations of Maximum Transconductance in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group



(a)



(b)

Figure A.5: Device-to-Device Variations of Gate Leakage Current in (a) the 60 Mrads Group and (b) the sequential 2 Mrads Group

To evaluate the general degradation evolution behavior of these EPC E-mode

GaN HEMTs, the mean variations of the DC parameters under all the characterized total doses, were calculated for all the irradiated devices, regardless of their group affiliations. The results of the calculations are plotted in Figure A.6-Figure A.8, and quantitatively summarized in Table A.1. The general variation trends of the DC parameters were consistent, with V_{th} , $g_{m,max}$ decreasing and $I_{g,leak}$ increasing. In agreement with the previous observations for devices in a group, for V_{th} , $g_{m,max}$, the beginning part of irradiation seemed to account for the majority of parametric variations, while at high doses (beyond a couple of Megarads) the degradations were much slower or even came to a halt. On the other hand, there was a more consistent increase for $I_{g,leak}$ throughout the 60 Mrads total dose. Nevertheless, for all the parameters there were some fluctuations between the adjacent characterization doses that countered the general variations trends. There was a clear influence of the device-to-device variations and possibly measurement errors, which may have a more significant impact given the small number of devices that were irradiated overall. Summarizing all the results from the irradiation test in the present work, it could be temporally concluded that a general behavior of device transfer characteristics degradation was the smaller the cumulative total doses the devices have sustained were, the higher the degradation rates were. On the other hand, for the gate leakage, the degradations were more consistent over a large range of irradiation doses.

Table A.1: Summary of mean variations of DC parameters for all groups of devices

Characterized Doses	Pre-rad	5 krad	20 krad	100 krad	600 krad	2 Mrads	10 Mrads	35 Mrads	60 Mrads
ΔV_{th}	0	-0.0824	-0.1280	-0.1150	-0.0753	-0.1655	-0.1436	-0.0919	-0.2209
Normalized $g_{m,max}$	1	0.7227	0.6512	0.6222	0.7206	0.5391	0.6896	0.5168	0.5737
Normalized $I_{g,leak}$	1	1.5394	1.6992	1.8680	1.1686	2.1544	5.0649	5.6499	27.6570

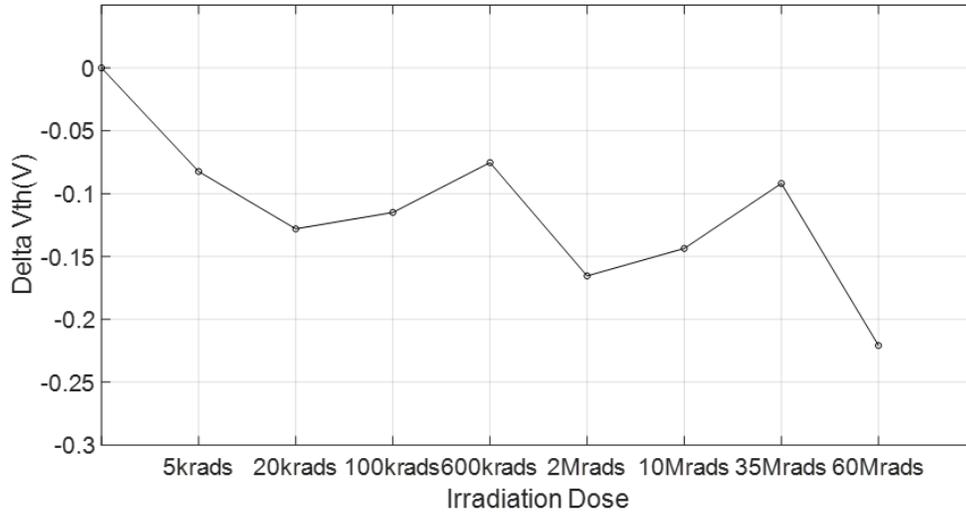


Figure A.6: Evolution of Mean ΔV_{th} Values for All Devices Irradiated and Characterized at All Doses

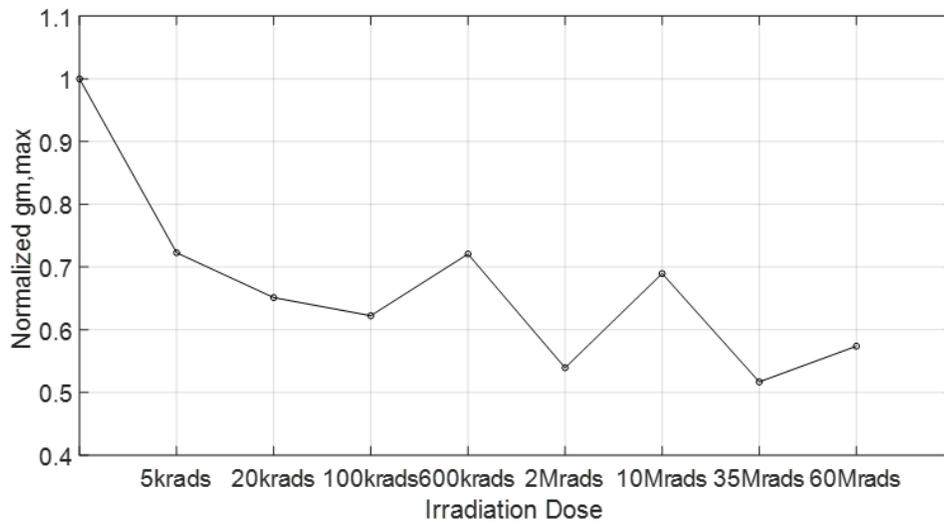


Figure A.7: Evolution of Mean Normalized $g_{m,max}$ Values for All Devices Irradiated and Characterized at All Doses

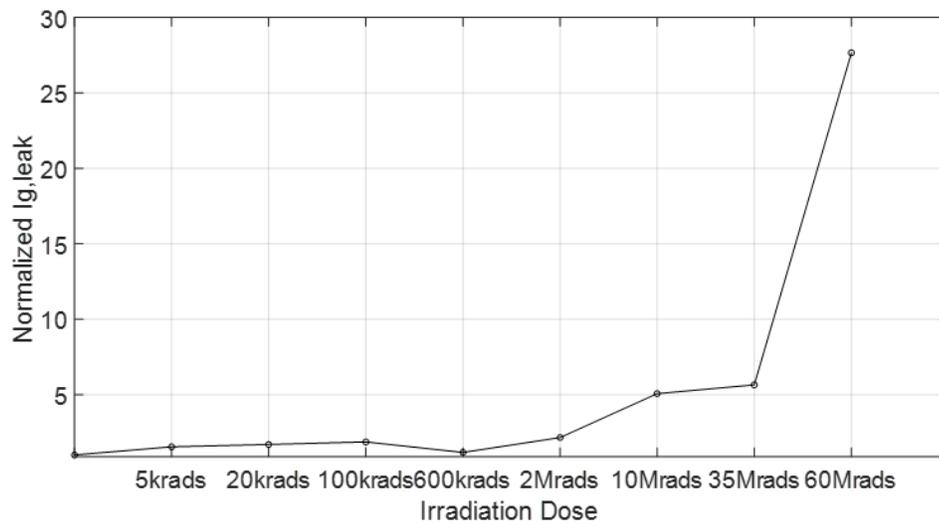


Figure A.8: Evolution of Mean Normalized $I_{g,leak}$ Values for All Devices Irradiated and Characterized at All Doses

A.6 Hypothesis of Degradation Mechanisms and Potential for Prognostic Model Implementation

The variations of V_{th} and $g_{m,max}$ could be attributed to defect generation throughout the AlGaIn and GaN layers and at the semiconductor/dielectric interfaces. In gamma radiation, high-energy photons produce 0.6 MeV Compton electrons, generating electron traps in the nitride layers, which are either shallow donor-like defects like nitrogen vacancies or the photo-generated holes [181]. As the irradiation dose increases, the created traps reduce the effective channel doping and carrier density, and the activation energies decrease thereby [182]. In the lower cumulative doses, many new positively-charged shallow electron traps are introduced, so an increasing number of electrons in the nitride layers are trapped, resulting in a decrease in the carrier concentration. However, since shallow traps also release the trapped electrons more easily, as the irradiation goes on, an increasing number of the trapped electrons are released, although new traps are introduced simultaneously trapping electrons. Therefore, at higher doses, the trapping rate and release rate of electrons in the nitride layers are not significantly different, almost leading to a ‘dynamic equilibrium’. In this case, even a high total dose of 50 Mrads would not greatly increase the total number of trapped electrons, explaining the minor degradation of the transfer characteristics at the second stage of the higher-dose irradiation.

As for the increase in $I_{g,leak}$, it was suggested in [183] that radiation damage of the barrier layer under the gate could account for the phenomenon. The crys-

talline imperfections in the AlGaIn layer increase significantly after the radiation, causing local variations in strain that result in charge redistribution around defects. With irradiation doses increasing, the number of imperfections continue to increase, accounting for the consistent increases in $I_{g,leak}$ at higher doses. Saturation of imperfections might show up at very high doses, but it may not be significant enough to put forward assessments based on the results in the present work.

Overall, the degradation mechanism of the electrical parameter variations still requires further investigations. It is hoped that the prognostic model developed in the present dissertation can be used in the radiation effect analysis. The determination of the exact mechanism will help find out an appropriate "driving variable" for the prognostic model. Also, more data instances are needed at lower radiation doses to enable densely sampled time-series degradation data of the DC parameters for anomaly detection and RUL prediction. It was discovered that during a 9.8-day space shuttle mission, the effective Gamma dose equivalent to the astronauts based on in-flight measurements was 4.1 mSv [184]. In 10 years, the total dose would be around 1500 mSv, or 150 rads. For electronic devices, given a smaller area of exposure compared with human bodies, the accumulated dose will be even smaller. Therefore, it is important to take more measurements at low doses of a few rads to a few krads to increase the fidelity of the parametric degradation procedure and enhance understanding of the degradation behavior at the low-dose range. After taking these steps, it will be possible to implement the prognostic model for Gamma-radiation-induced degradation at low doses.

A.7 Summary

This appendix is reporting on gamma radiation on some commercially available EPC2035 E-mode GaN HEMTs with a variety of total doses ranging from 5 krads to 60 Mrads, to investigate their DC parameter degradations in hard radiation environments. A higher-dose irradiation with two stages of up to 60 Mrads in total was conducted on 8 devices. A lower-dose irradiation was performed on another 6 devices with a total dose of up to 2 Mrads the devices. For the transfer characteristics, significant degradations were gauged within the first couple of Megarads, while much smaller amount of degradation was observed for further doses. The gate leakage, however, continuously intensified with accumulated total dose. These results indicated that the devices were not radiation resistant and degradations were already significant within several kilorads of gamma radiation. The degradation mechanism is believed to be clustering of the radiation-induced shallow electron traps that under higher cumulative dose, exhibits a dynamic equilibrium of trapping and release of electrons.

Bibliography

- [1] T Paul Chow. High-voltage sic and gan power devices. *Microelectronic Engineering*, 83(1):112–122, 2006.
- [2] Lotfi Saidi, Jaouher Ben Ali, Mohamed Benbouzid, and Eric Bechhofer. An integrated wind turbine failures prognostic approach implementing kalman smoother with confidence bounds. *Applied Acoustics*, 138:199–208, 2018.
- [3] Frede Blaabjerg and Zhe Chen. Power electronics for modern wind turbines. *Synthesis Lectures on Power Electronics*, 1(1):1–68, 2005.
- [4] Remus Teodorescu, Marco Liserre, and Pedro Rodríguez. Grid converter structures for wind turbine systems. *Grid Converters for Photovoltaic and Wind Power Systems*, pages 123–143, 2011.
- [5] Infineon Technologies AG. *EasyPACK module with Trench/Fieldstop IGBT3 and Emitter Controlled 3 diode and NTC*, 10 2013. Rev. 2.1.
- [6] Katharina Fischer, Thomas Stalin, Hans Ramberg, T Thiringer, Jan Wenske, and Robert Karlsson. Investigation of converter failure in wind turbines. Technical report, 11 2012.
- [7] He Xiao-Guang, Zhao De-Gang, and Jiang De-Sheng. Formation of two-dimensional electron gas at algan/gan heterostructure and the derivation of its sheet density expression. *Chinese Physics B*, 24(6):067301, 2015.
- [8] M Tounsi, Amrane Oukaour, Boubekour Tala-Ighil, Hamid Gualous, Bertrand Boudart, and Djamil Aissani. Characterization of high-voltage igt module degradations under pwm power cycling test at high ambient temperature. *Microelectronics Reliability*, 50(9-11):1810–1814, 2010.
- [9] Tiancheng Li, Miodrag Bolic, and Petar M Djuric. Resampling methods for particle filtering: classification, implementation, and strategies. *IEEE Signal Processing Magazine*, 32(3):70–86, 2015.

- [10] Ilker Yildirim. Bayesian inference: metropolis-hastings sampling. *Department of Brain and Cognitive Sciences, University of Rochester, Rochester*, 2012.
- [11] Uwe Scheuermann and Ralf Schmidt. Impact of solder fatigue on module lifetime in power cycling tests. In *Power Electronics and Applications (EPE 2011), Proceedings of the 2011-14th European Conference on*, pages 1–10. IEEE, 2011.
- [12] R. Schmidt, F. Zeyss, and U. Scheuermann. Impact of absolute junction temperature on power cycling lifetime. In *2013 15th European Conference on Power Electronics and Applications (EPE)*, pages 1–10, Sept 2013.
- [13] S Narumanchi, K Bennion, D DeVoto, G Moreno, J Rugh, and S Waye. Advanced power electronics and electric motors annual report–2013. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.
- [14] Burak Ozpineci. Oak ridge national laboratory annual progress report for the electric drive technologies program. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). Power Electronics and Electric Machinery Research Facility, 2015.
- [15] Shaoyong Yang, Angus Bryant, Philip Mawby, Dawei Xiang, Li Ran, and Peter Tavner. An industry-based survey of reliability in power electronic converters. *IEEE transactions on Industry Applications*, 47(3):1441–1451, 2011.
- [16] Michael Pecht and Rubyca Jaai. A prognostics and health management roadmap for information and electronics-rich systems. *Microelectronics Reliability*, 50(3):317–323, 2010.
- [17] Andrew K.S. Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483 – 1510, 2006.
- [18] Jay Lee, Fangji Wu, Wenyu Zhao, Masoud Ghaffari, Linxia Liao, and David Siegel. Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1):314 – 334, 2014.
- [19] A. Kabir, C. Bailey, H. Lu, and S. Stoyanov. A review of data-driven prognostics in power electronics. In *2012 35th International Spring Seminar on Electronics Technology*, pages 189–192, May 2012.
- [20] URHO Pulkkinen. A stochastic model for wear prediction through condition monitoring. *Operational reliability and systematic maintenance*, pages 223–243, 1991.
- [21] Asok Ray and Sekhar Tangirala. A nonlinear stochastic model of fatigue crack dynamics. *Probabilistic Engineering Mechanics*, 12(1):33–40, 1997.

- [22] RW Swindeman and MJ Swindeman. A comparison of creep models for nickel base alloys for advanced energy systems. *International Journal of Pressure Vessels and Piping*, 85(1-2):72–79, 2008.
- [23] Michael Pecht. *Prognostics and health management of electronics*. Wiley Online Library, 2008.
- [24] Matthew J Daigle and Kai Goebel. Model-based prognostics with concurrent damage progression processes. *IEEE Transactions on Systems, man, and cybernetics: systems*, 43(3):535–546, 2013.
- [25] C. Bailey, C. Yin, H. Lu, M. Musallam, and C. M. Johnson. Current status of prognostics techniques and application to power electronics. In *2010 6th International Conference on Integrated Power Electronics Systems*, pages 1–6, March 2010.
- [26] Jianhui Luo, M. Namburu, K. Pattipati, Liu Qiao, M. Kawamoto, and S. Chigusa. Model-based prognostic techniques [maintenance applications]. In *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference.*, pages 330–340, Sept 2003.
- [27] Mark Schwabacher. A survey of data-driven prognostics. In *Infotech@ Aerospace*, page 7002. 2005.
- [28] M. Thirumarimurugan, N. Bagyalakshmi, and P. Paarkavi. Comparison of fault detection and isolation methods: A review. In *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, pages 1–6, Jan 2016.
- [29] Marcos E Orchard and George J Vachtsevanos. A particle filtering-based framework for real-time fault diagnosis and failure prognosis in a turbine engine. In *Control & Automation, 2007. MED'07. Mediterranean Conference on*, pages 1–6. IEEE, 2007.
- [30] Amar Khoukhi, H Khalid, R Doraiswami, and L Cheded. Fault detection and classification using kalman filter and hybrid neuro-fuzzy systems. *International Journal of Computer Applications*, 45(22):7–14, 2012.
- [31] Romano Patrick-Aldaco. *A model based framework for fault diagnosis and prognosis of dynamical systems with an application to helicopter transmissions*. PhD thesis, Georgia Institute of Technology, 2007.
- [32] Steven X Ding. *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.
- [33] Kai Goebel, Bhaskar Saha, and Abhinav Saxena. A comparison of three data-driven techniques for prognostics. In *62nd meeting of the society for machinery failure prevention technology (mfpt)*, pages 119–131, 2008.

- [34] Eric WT Ngai, Li Xiu, and Dorothy CK Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602, 2009.
- [35] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [36] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [37] Geetha Jagannathan, Claire Monteleoni, and Krishnan Pillaipakkamnatt. A semi-supervised learning approach to differential privacy. In *2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW)*, pages 841–848. IEEE, 2013.
- [38] Rana Aamir Raza Ashfaq, Xi-Zhao Wang, Joshua Zhexue Huang, Haider Abbas, and Yu-Lin He. Fuzziness based semi-supervised learning approach for intrusion detection system. *Information Sciences*, 378:484–497, 2017.
- [39] Ye Zhao, Roy Ball, Jerry Mosesian, Jean-François de Palma, and Brad Lehman. Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays. *IEEE Transactions on Power Electronics*, 30(5):2848–2858, 2015.
- [40] Rowland R Sillito and Robert B Fisher. Semi-supervised learning for anomalous trajectory detection. In *BMVC*, volume 1, pages 035–1, 2008.
- [41] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [42] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos.
- [43] Aleksandar Lazarevic, Arindam Banerjee, Varun Chandola, Vipin Kumar, and Jaideep Srivastava. Data mining for anomaly detection.
- [44] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on signal processing*, 50(2):174–188, 2002.
- [45] Simon Maskell and Neil Gordon. A tutorial on particle filters for on-line nonlinear/non-gaussian bayesian tracking. *Target tracking: algorithms and applications (Ref. No. 2001/174)*, *IEE*, 2:21–215, 2001.
- [46] Arnaud Doucet and Adam M Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

- [47] Bhaskar Saha, Kai Goebel, and Jon Christophersen. Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, 31(3-4):293–308, 2009.
- [48] Jose Celaya, Abhinav Saxena, Sankalita Saha, and Kai F Goebel. Prognostics of power mosfets under thermal stress accelerated aging using data-driven and model-based methodologies. 2011.
- [49] Bhaskar Saha, Jose R Celaya, Philip F Wysocki, and Kai F Goebel. Towards prognostics for electronics components. In *Aerospace conference, 2009 IEEE*, pages 1–7. IEEE, 2009.
- [50] Nishad Patil, Diganta Das, and Michael Pecht. A prognostic approach for non-punch through and field stop igbts. *Microelectronics Reliability*, 52(3):482–488, 2012.
- [51] Moinul Shahidul Haque, Seungdeog Choi, and Jeihoon Baek. Auxiliary particle filtering-based estimation of remaining useful life of igbt. *IEEE Transactions on Industrial Electronics*, 65(3):2693–2703, 2018.
- [52] Marco Rigamonti, Piero Baraldi, Enrico Zio, Daniel Astigarraga, and Ainhoa Galarza. Particle filter-based prognostics for an electrolytic capacitor working in variable operating conditions. *IEEE Transactions on Power Electronics*, 31(2):1567–1575, 2016.
- [53] Lifeng Wu, Jing Yang, Zhen Peng, and Hongmin Wang. Remaining useful life prognostic of power metal oxide semiconductor field effect transistor based on improved particle filter algorithm. *Advances in Mechanical Engineering*, 9(12):1687814017749324, 2017.
- [54] Cen Chen, Xuerong Ye, Yixing Wang, Junzhong Xu, and Guofu Zhai. Phm application of power converters using health precursor of power mosfets. In *Prognostics and System Health Management Conference (PHM), 2015*, pages 1–5. IEEE, 2015.
- [55] Cen Chen, Xuerong Ye, Han Wang, Guofu Zhai, and Ran Wan. In-situ prognostic method of power mosfet based on miller effect. In *Prognostics and System Health Management Conference (PHM-Harbin), 2017*, pages 1–5. IEEE, 2017.
- [56] Daeil Kwon and Jeongah Yoon. A model-based prognostic approach to predict interconnect failure using impedance analysis. *Journal of Mechanical Science and Technology*, 30(10):4447–4452, 2016.
- [57] Robert W Erickson and Dragan Maksimovic. *Fundamentals of power electronics*. Springer Science & Business Media, 2007.
- [58] Leif Halbo and Per Ohlckers. *Electronic components, packaging and production*. Oslo Universitet, 1995.

- [59] Shaoyong Yang, Dawei Xiang, Angus Bryant, Philip Mawby, Li Ran, and Peter Tavner. Condition monitoring for device reliability in power electronic converters: A review. *IEEE Transactions on Power Electronics*, 25(11):2734–2752, 2010.
- [60] E Wolfgang. Examples for failures in power electronics systems. *ECPE tutorial on reliability of power electronic systems, Nuremberg, Germany*, pages 19–20, 2007.
- [61] Michael Wilkinson, Keir Harman, Ben Hendriks, Fabio Spinato, Thomas van Delft, GL Garrad, and UK Thomas. Measuring wind turbine reliability, results of the reliawind project. In *EWEA Conference*, pages 1–8, 2011.
- [62] Vinod Kumar Khanna. *Insulated gate bipolar transistor IGBT theory and design*. John Wiley & Sons, 2004.
- [63] B Jayant Baliga. Enhancement-and depletion-mode vertical-channel mos gated thyristors. *Electronics Letters*, 15(20):645–647, 1979.
- [64] Jayant Baliga. Igbt: The ge story [a look back]. *IEEE Power Electronics Magazine*, 2(2):16–23, 2015.
- [65] Abdus Sattar et al. Insulated gate bipolar transistor (igbt) basics. In *IXYS Corporation. IXAN0063*. Citeseer, 1998.
- [66] Michael Pecht, Rakish Agarwal, F Patrick McCluskey, Terrance J Dishongh, Sirius Javadpour, and Rahul Mahajan. *Electronic packaging materials and their properties*. CRC press, 1998.
- [67] M. Boettcher and F. W. Fuchs. Power electronic converters in wind energy systems 2014; considerations of reliability and strategies for increasing availability. In *Proceedings of the 2011 14th European Conference on Power Electronics and Applications*, pages 1–10, Aug 2011.
- [68] H. Wang, M. Liserre, F. Blaabjerg, P. de Place Rimmen, J. B. Jacobsen, T. Kvisgaard, and J. Landkildehus. Transitioning to physics-of-failure as a reliability driver in power electronics. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2(1):97–114, 2014.
- [69] Donald A Gajewski, Scott Sheppard, Simon Wood, Jeff B Barner, Jim Milligan, and John Palmour. Reliability comparison of 28 v–50 v gan-on-sic s-band and x-band technologies.
- [70] M. Held, P. Jacob, G. Nicoletti, P. Scacco, and M. H. Poech. Fast power cycling test of igbt modules in traction application. In *Proceedings of Second International Conference on Power Electronics and Drive Systems*, volume 1, pages 425–430 vol.1, May 1997.

- [71] Mauro Ciappa. Selected failure mechanisms of modern power modules. *Microelectronics reliability*, 42(4-5):653–667, 2002.
- [72] E Herr, T Frey, Reinhard Schlegel, A Stuck, and R Zehringer. Substrate-to-base solder joint reliability in high power igbt modules. 37:1719–1722, 10 1997.
- [73] Wayne Nelson. Accelerated life testing-step-stress models and data analyses. *IEEE transactions on reliability*, 29(2):103–108, 1980.
- [74] Mohammad Modarres, Mehdi Amiri, and Christopher Jackson. *Probabilistic Physics of Failure Approach to Reliability: Modeling, Accelerated Testing, Prognosis and Reliability Assessment*. John Wiley & Sons, 2017.
- [75] Jeffrey W Harms. Revision of mil-hdbk-217, reliability prediction of electronic equipment. In *Reliability and Maintainability Symposium (RAMS), 2010 Proceedings-Annual*, pages 1–3. IEEE, 2010.
- [76] Mark White. Microelectronics reliability: physics-of-failure based modeling and lifetime evaluation. Technical report, Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2008., 2008.
- [77] JEDEC Solid State Technology Association. Jedec standard-power cycling, 2007.
- [78] Josef Lutz, Heinrich Schlangenotto, Uwe Scheuermann, and Rik De Doncker. *Semiconductor power devices: physics, characteristics, reliability*. Springer Science & Business Media, 2011.
- [79] Robert W Keyes. Physical limits of silicon transistors and circuits. *Reports on Progress in Physics*, 68(12):2701, 2005.
- [80] Kiyoshi Takahashi, Akihiko Yoshikawa, and Adarsh Sandhu. *Wide bandgap semiconductors*. Springer-Verlag Berlin Heidelberg., 2007.
- [81] M Sawant and A Christou. Failure modes and effects criticality analysis and accelerated life testing of leds for medical applications. *Solid-State Electronics*, 78:39–45, 2012.
- [82] Matteo Meneghini, Oliver Hilt, Joachim Wuerfl, and Gaudenzio Meneghesso. Technology and reliability of normally-off gan hemts with p-type gate. *Energies*, 10(2):153, 2017.
- [83] Gaudenzio Meneghesso, Giovanni Verzellesi, Francesca Danesin, Fabiana Rampazzo, Franco Zanon, Augusto Tazzoli, Matteo Meneghini, and Enrico Zanoni. Reliability of gan high-electron-mobility transistors: State of the art and perspectives. *IEEE Transactions on Device and Materials Reliability*, 8(2):332–343, 2008.

- [84] Takashi Mimura. The early history of the high electron mobility transistor (hemt). *IEEE Transactions on microwave theory and techniques*, 50(3):780–782, 2002.
- [85] David L Pulfrey. *Understanding modern transistors and diodes*. Cambridge University Press, 2010.
- [86] Dag Grini. Rf basics, rf for non-rf engineers. 2006.
- [87] F Fornetti, KA Morris, and MA Beach. Pulsed operation and performance of commercial gan hemts. In *Microwave Integrated Circuits Conference, 2009. EuMIC 2009. European*, pages 226–229. IEEE, 2009.
- [88] Umesh K Mishra, Primit Parikh, and Yi-Feng Wu. Algan/gan hemts-an overview of device operation and applications. *Proceedings of the IEEE*, 90(6):1022–1031, 2002.
- [89] Enrico Zanoni. Gan hemt reliability research—a white paper—. 2017.
- [90] V. Tilak, B. Green, H. Kim, R. Dimitrov, J. Smart, W. J. Schaff, J. R. Shealy, and L. F. Eastman. Effect of passivation on algan/gan hemt device performance. In *2000 IEEE International Symposium on Compound Semiconductors. Proceedings of the IEEE Twenty-Seventh International Symposium on Compound Semiconductors (Cat. No.00TH8498)*, pages 357–363, Oct 2000.
- [91] Jungwoo Joh and Jesus A del Alamo. Mechanisms for electrical degradation of gan high-electron mobility transistors. In *Electron Devices Meeting, 2006. IEDM'06. International*, pages 1–4. IEEE, 2006.
- [92] Y. Chen, R. Coffie, W. Luo, M. Wojtowicz, I. Smorchkova, B. Heying, Y. Kim, M. V. Aust, and A. Oki. Survivability of algan/gan hemt. In *2007 IEEE/MTT-S International Microwave Symposium*, pages 307–310, June 2007.
- [93] R. Coffie, Y. Chen, I. P. Smorchkova, B. Heying, V. Gambin, W. Sutton, Y. . Chou, W. . Luo, M. Wojtowicz, and A. Oki. Temperature and voltage dependent rf degradation study in algan/gan hemts. In *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, pages 568–569, April 2007.
- [94] P. Valizadeh and D. Pavlidis. Effects of rf and dc stress on algan/gan modfets: a low-frequency noise-based investigation. *IEEE Transactions on Device and Materials Reliability*, 5(3):555–563, Sept 2005.
- [95] A. Sasikumar, A.R. Arehart, G.D. Via, B. Wunningham, B. Poling, E. Heller, and S.A. Ringel. Identification of an rf degradation mechanism in gan based hemts triggered by midgap traps. *Microelectronics Reliability*, 55(11):2258 – 2262, 2015.

- [96] A.R. Arehart, A. Sasikumar, G.D. Via, B. Poling, E.R. Heller, and S.A. Ringel. Evidence for causality between gan rf hemt degradation and the ec-0.57ev trap in gan. *Microelectronics Reliability*, 56:45 – 48, 2016.
- [97] Gokhan Kurt, Ahmet Toprak, Ozlem A Sen, and Ekmel Ozbay. Effect of field plate length on power performance of gan based hemts. In *Proceedings of the 18th International Conference On Circuits (Part of CSCC'14), Advances Robotics, Mechatronics and Circuits*, 2014.
- [98] AM Conway, M Chen, P Hashimoto, PJ Willadsen, and M Micovic. Accelerated rf life testing of gan hfets. In *Reliability physics symposium, 2007. proceedings. 45th annual. ieee international*, pages 472–475. IEEE, 2007.
- [99] Rambabu Kandepu, Bjarne Foss, and Lars Imsland. Applying the unscented kalman filter for nonlinear state estimation. *Journal of process control*, 18(7-8):753–768, 2008.
- [100] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F - Radar and Signal Processing*, 140(2):107–113, April 1993.
- [101] Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.
- [102] N. Oudjane and C. Musso. Progressive correction for regularized particle filters. In *Proceedings of the Third International Conference on Information Fusion*, volume 2, pages THB2/10–THB2/17 vol.2, July 2000.
- [103] Marine Jouin, Rafael Gouriveau, Daniel Hissel, Marie-Cécile Péra, and Nouredine Zerhouni. Particle filter-based prognostics: Review, discussion and perspectives. *Mechanical Systems and Signal Processing*, 72:2–31, 2016.
- [104] Qiang Miao, Lei Xie, Hengjuan Cui, Wei Liang, and Michael Pecht. Remaining useful life prediction of lithium-ion battery with unscented particle filter technique. *Microelectronics Reliability*, 53(6):805–810, 2013.
- [105] A. Ahmed, S. S. Islam, and A. F. M. Anwar. Gain compression in gan hemt amplifiers. In *2001 International Semiconductor Device Research Symposium. Symposium Proceedings (Cat. No.01EX497)*, pages 205–208, Dec 2001.
- [106] Hemant P Rao and Gijs Bosman. Study of rf reliability of gan hemts using low-frequency noise spectroscopy. *IEEE Transactions on Device and Materials Reliability*, 12(1):31–36, 2012.
- [107] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, 2016.
- [108] James Cannady. Artificial neural networks for misuse detection. In *National information systems security conference*, volume 26. Baltimore, 1998.

- [109] Imola K Fodor. A survey of dimension reduction techniques. *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*, 9:1–18, 2002.
- [110] Stan Z. Li and Anil Jain, editors. *Mahalanobis Distance*, pages 953–953. Springer US, Boston, MA, 2009.
- [111] R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1 – 18, 2000.
- [112] Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.
- [113] Sami Äyrämö and Tommi Kärkkäinen. Introduction to partitioning-based clustering methods with a robust example. *Reports of the Department of Mathematical Information Technology. Series C, Software engineering and computational intelligence*, (1/2006), 2006.
- [114] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.
- [115] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.
- [116] Patrick O’Connor and Andre Kleyner. *Practical reliability engineering*. John Wiley & Sons, 2012.
- [117] T. T. T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys Tutorials*, 10(4):56–76, Fourth 2008.
- [118] Yuhui Zhao, Suxia Yu, Bingbing Chu, Nan Zhang, and Xin Hu. Classification of three wine varieties based on elm and pca. In *International Conference on Intelligent Science and Intelligent Data Engineering*, pages 647–654. Springer, 2012.
- [119] Saket Sathe and Charu Aggarwal. Lodes: local density meets spectral outlier detection. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 171–179. SIAM, 2016.
- [120] Lei Nie, Michael H Azarian, Mohammadreza Keimasi, and Michael Pecht. Prognostics of ceramic capacitor temperature-humidity-bias reliability using mahalanobis distance analysis. *Circuit World*, 33(3):21–28, 2007.
- [121] William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.
- [122] Ka Yee Yeung and Walter L Ruzzo. Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. 2001.

- [123] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, Dec 1985.
- [124] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(Oct):2837–2854, 2010.
- [125] Bhaskar Saha and Kai Goebel. Uncertainty management for diagnostics and prognostics of batteries using bayesian techniques. In *Aerospace Conference, 2008 IEEE*, pages 1–8. IEEE, 2008.
- [126] Chaochao Chen and Michael Pecht. Prognostics of lithium-ion batteries using model-based and data-driven methods. In *Prognostics and System Health Management (PHM), 2012 IEEE Conference on*, pages 1–6. IEEE, 2012.
- [127] Royce G Forman, V Shivakumar, and James C Newman Jr. Fatigue-crack-growth computer program. 1991.
- [128] Chunyan Jiang, Ting Liu, Chunhua du, Xin Huang, Mengmeng Liu, Zhao Zhenfu, Linxuan Li, Xiong Pu, Junyi Zhai, Weiguo Hu, and Zhong Wang. Piezotronic effect tuned algan/gan high electron mobility transistor. *Nanotechnology*, 28, 09 2017.
- [129] Donald A Gajewski, Scott Sheppard, Tina McNulty, Jeff B Barner, Jim Milligan, and John Palmour. Reliability of gan/algan hemt mmic technology on 100-mm 4h-sic. In *Reliability of Compound Semiconductors Workshop*, 2011.
- [130] Raymond S Pengelly, Simon M Wood, James W Milligan, Scott T Sheppard, and William L Pribble. A review of gan on sic high electron-mobility power transistors and mmics. *IEEE Transactions on Microwave Theory and Techniques*, 60(6):1764–1783, 2012.
- [131] N. Komai, M. Watanabe, and Z. Horita. Interdiffusivity measurements and interface observations using ni/ni₃ge diffusion couples. *Acta Metallurgica et Materialia*, 43(8):2967 – 2974, 1995.
- [132] TR Paul, IV Belova, EV Levchenko, AV Evteev, and GE Murch. Determining a tracer diffusivity by way of the darken-manning equation for interdiffusion in binary alloy systems. In *Diffusion Foundations*, volume 4, pages 25–54. Trans Tech Publ, 2015.
- [133] Ursula R Kattner and Carelyn E Campbell. Invited review: Modelling of thermodynamics and diffusion in multicomponent systems. *Materials Science and Technology*, 25(4):443–459, 2009.
- [134] Richard JD Tilley. *Principles and applications of chemical defects*. CRC Press, 1998.

- [135] Devendra Gupta. *Diffusion processes in advanced technological materials*. Springer Science & Business Media, 2010.
- [136] William J Reed. Power-law adjusted survival models. *Communications in Statistics-Theory and Methods*, 41(20):3692–3703, 2012.
- [137] David E Rupp and Ross A Woods. Increased flexibility in base flow modelling using a power law transmissivity profile. *Hydrological Processes: An International Journal*, 22(14):2667–2671, 2008.
- [138] Matteo Corbetta, S Sbarufatti, and Marco Giglio. Optimal tuning of particle filtering random noise for monotonic degradation processes. In *Proceedings of the third European Conference of the Prognostics and Health Management Society*, 2016.
- [139] Jeroen D Hol. Resampling in particle filters, 2004.
- [140] Edward R Beadle and Petar M Djuric. A fast-weighted bayesian bootstrap filter for nonlinear model state estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 33(1):338–343, 1997.
- [141] Miodrag Bolić, Petar M Djurić, and Sangjin Hong. Resampling algorithms for particle filters: A computational complexity perspective. *EURASIP Journal on Advances in Signal Processing*, 2004(15):403686, 2004.
- [142] Miodrag Bolic, Petar M Djuric, and Sangjin Hong. New resampling algorithms for particle filters. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 2, pages II–589. IEEE, 2003.
- [143] Dan Crisan and Terry Lyons. A particle approximation of the solution of the kushner–stratonovitch equation. *Probability Theory and Related Fields*, 115(4):549–578, 1999.
- [144] Tiancheng Li, Tariq P Sattar, and Dedong Tang. A fast resampling scheme for particle filters. 2013.
- [145] Jun S Liu, Rong Chen, and Tanya Logvinenko. A theoretical framework for sequential importance sampling with resampling. In *Sequential Monte Carlo methods in practice*, pages 225–246. Springer, 2001.
- [146] Tiancheng Li, Shudong Sun, and Tariq Pervez Sattar. Adapting sample size in particle filters through kld-resampling. *Electronics Letters*, 49(12):740–742, 2013.
- [147] BG Sileshi, Carles Ferrer, and Juan Oliver. Particle filters and resampling techniques: Importance in computational complexity analysis. In *Design and Architectures for Signal and Image Processing (DASIP), 2013 Conference on*, pages 319–325. IEEE, 2013.

- [148] Marco Fraccaro, Ulrich Paquet, and Ole Winther. An adaptive resample-move algorithm for estimating normalizing constants. *arXiv preprint arXiv:1604.01972*, 2016.
- [149] E. Wilson. *Gamma Rays: Study of Electromagnetic Radiation*. NY Research Press, 2015.
- [150] Ndt resource center, October 11, 2007 2001.
- [151] Gordon Gilmore. *Practical gamma-ray spectroscopy*. John Wiley & Sons, 2011.
- [152] L.A. Franks and Society of Photo-optical Instrumentation Engineers. *Hard X-ray and Gamma-ray Detector Physics and Penetrating Radiation Systems VIII: 14-17 August, 2006, San Diego, California, USA*. Number 6319. SPIE, 2006.
- [153] Kevin J Chen and Chunhua Zhou. Enhancement-mode algan/gan hemt and mis-hemt technology. *physica status solidi (a)*, 208(2):434–438, 2011.
- [154] Sumit Verma, Sajad A. Loan, and Abdullah G. Alharbi. Polarization engineered enhancement mode gan hemt: Design and investigation. *Superlattices and Microstructures*, 119:181 – 193, 2018.
- [155] WB Lanford, T Tanaka, Y Otoki, and I Adesida. Recessed-gate enhancement-mode gan hemt with high threshold voltage. *Electronics Letters*, 41(7):449–450, 2005.
- [156] Ki-Sik Im, Jong-Bong Ha, Ki-Won Kim, Jong-Sub Lee, Dong-Seok Kim, Sung-Ho Hahm, and Jung-Hee Lee. Normally off gan mosfet based on algan/gan heterostructure with extremely high 2deg density grown on silicon substrate. *IEEE Electron Device Letters*, 31(3):192–194, 2010.
- [157] Jungwoo Joh, Feng Gao, Tomás Palacios, and Jesús A. del Alamo. A model for the critical voltage for electrical degradation of gan high electron mobility transistors. *Microelectronics Reliability*, 50(6):767 – 773, 2010. 2009 Reliability of Compound Semiconductors (ROCS) Workshop).
- [158] W. Chen, K. Wong, and K. J. Chen. Monolithic integration of lateral field-effect rectifier with normally-off hemt for gan-on-si switch-mode power supply converters. In *2008 IEEE International Electron Devices Meeting*, pages 1–4, Dec 2008.
- [159] Yong Cai, Yugang Zhou, K. J. Chen, and K. M. Lau. High-performance enhancement-mode algan/gan hemts using fluoride-based plasma treatment. *IEEE Electron Device Letters*, 26(7):435–437, July 2005.
- [160] M. Kanamura, T. Ohki, T. Kikkawa, K. Imanishi, T. Imada, A. Yamada, and N. Hara. Enhancement-mode gan mis-hemts with n-gan/i-aln/n-gan triple cap layer and high- k gate dielectrics. *IEEE Electron Device Letters*, 31(3):189–191, March 2010.

- [161] S. L. Selvaraj, K. Nagai, and T. Egawa. MOCVD grown normally-off type AlGaN/GaN HEMTs on 4 inch Si using p-AlGaN cap layer with high breakdown. In *68th Device Research Conference*, pages 135–136, June 2010.
- [162] Injun Hwang, Hyoji Choi, JaeWon Lee, Hyuk Soon Choi, Jongseob Kim, Jongbong Ha, Chang-Yong Um, Sun-Kyu Hwang, Jaejoon Oh, Jun-Youn Kim, et al. 1.6 kV, 2.9 mW/cm² normally-off p-GaN HEMT device. In *Power Semiconductor Devices and ICs (ISPSD), 2012 24th International Symposium on*, pages 41–44. IEEE, 2012.
- [163] K. Ota, K. Endo, Y. Okamoto, Y. Ando, H. Miyamoto, and H. Shimawaki. A normally-off GaN FET with high threshold voltage uniformity using a novel piezo neutralization technique. In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4, Dec 2009.
- [164] M Ishida, M Kuroda, T Ueda, and T Tanaka. Nonpolar AlGaN/GaN HEMTs with a normally off operation. *Semiconductor Science and Technology*, 27(2):024019, 2012.
- [165] H. Jiang, C. W. Tang, and K. M. Lau. Enhancement-mode GaN MOS-HEMTs with recess-free barrier engineering and high- k ZnO gate dielectric. *IEEE Electron Device Letters*, 39(3):405–408, March 2018.
- [166] SA Goodman, FD Auret, MJ Legodi, B Beaumont, and P Gibart. Characterization of electron-irradiated n-GaN. *Applied Physics Letters*, 78(24):3815–3817, 2001.
- [167] Nathan E Ives, Jin Chen, Arthur F Witulski, Ronald D Schrimpf, Daniel M Fleetwood, Ralph W Bruce, Michael W McCurdy, En Xia Zhang, and Lloyd W Massengill. Effects of proton-induced displacement damage on gallium nitride HEMTs in RF power amplifier applications. *IEEE Transactions on Nuclear Science*, 62(6):2417–2422, 2015.
- [168] Jin Chen, Yevgeniy S Puzyrev, Rong Jiang, En Xia Zhang, Michael W McCurdy, Daniel M Fleetwood, Ronald D Schrimpf, Sokrates T Pantelides, Aaron R Arehart, Steven A Ringel, et al. Effects of applied bias and high field stress on the radiation response of GaN/AlGaN HEMTs. *IEEE Transactions on Nuclear Science*, 62(6):2423–2430, 2015.
- [169] YS Puzyrev, T Roy, EX Zhang, DM Fleetwood, RD Schrimpf, and ST Pantelides. Radiation-induced defect evolution and electrical degradation of AlGaN/GaN high-electron-mobility transistors. *IEEE Transactions on Nuclear Science*, 58(6):2918–2924, 2011.
- [170] Lü Ling, Zhang Jin-Cheng, Xue Jun-Shuai, Ma Xiao-Hua, Zhang Wei, Bi Zhi-Wei, Zhang Yue, and Hao Yue. Neutron irradiation effects on AlGaN/GaN high electron mobility transistors. *Chinese physics B*, 21(3):037104, 2012.

- [171] Rong Jiang, En Xia Zhang, Xiao Shen, Jin Chen, Kai Ni, Pan Wang, Daniel M Fleetwood, Ronald D Schrimpf, Stephen W Kaun, Erin CH Kyle, et al. Total ionizing dose effects in passivated and unpassivated algan/gan hemts. In *Radiation and Its Effects on Components and Systems (RADECS), 2016 16th European Conference on*, pages 1–4. IEEE, 2016.
- [172] Sorina Lazanu, Ionel Lazanu, Emilio Borch, and Mara Bruzzi. Theoretical calculations of the primary defects induced by pions and protons in sic. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 485(3):768–773, 2002.
- [173] Geoffrey P Summers, Edward A Burke, Philip Shapiro, Scott R Messenger, and Robert J Walters. Damage correlations in semiconductors exposed to gamma, electron and proton radiations. *IEEE Transactions on Nuclear Science*, 40(6):1372–1379, 1993.
- [174] O Aktas, A Kuliev, V Kumar, R Schwindt, S Toshkov, D Costescu, J Stubbins, and I Adesida. 60co gamma radiation effects on dc, rf, and pulsed i–v characteristics of algan/gan hemts. *Solid-State Electronics*, 48(3):471–475, 2004.
- [175] A. E. Belyaev, R. V. Konakova, M. V. Petrychuk, A. Y. Avksentyev, A. M. Kurakin, S. A. Vitusevich, N. Klein, S. V. Danylyuk, and B. A. Danylchenko. The effect of gamma-irradiation on the operating parameters of group iii nitrides-based field effect transistors. In *12th International Conference Microwave and Telecommunication Technology*, pages 484–485, Sept 2002.
- [176] Hong-Yeol Kim, Jihyun Kim, Lu Liu, Chien-Fong Lo, Fan Ren, and Stephen J Pearton. Electrical characterization of 60co gamma radiation-exposed inaln/gan high electron mobility transistors. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, 31(5):051210, 2013.
- [177] Travis J Anderson, Andrew D Koehler, Marko J Tadjer, Karl D Hobart, Petra Specht, Matthew Porter, Todd R Weatherford, Brad Weaver, Jennifer K Hite, and Fritz J Kub. Reliability of gan hemts: Electrical and radiation-induced failure mechanism. *ECS Transactions*, 58(4):221–225, 2013.
- [178] MD Smith, D O’Mahony, F Vitobello, M Muschitiello, A Costantino, AR Barnes, and PJ Parbrook. A comparison of the 60co gamma radiation hardness, breakdown characteristics and the effect of sinx capping on inaln and algan hemts for space applications. *Semiconductor Science and Technology*, 31(2):025008, 2015.
- [179] EPC Corporation. *Enhancement Mode Power Transistor*, 6 2018. Rev. June 2018.

- [180] Mark Pavier, Arthur Woodworth, Andrew Sawle, Hurst Green, Ralph Monteiro, Carl Blake, and Jason Chiu-El. Understanding the effect of power mosfet package parasitics on vrm circuit efficiency at frequencies above 1mhz. 11 2018.
- [181] GA Umana-Membreno, JM Dell, TP Hessler, BD Nener, G Parish, L Faraone, and UK Mishra. 60 co gamma-irradiation-induced defects in n-gan. *Applied physics letters*, 80(23):4354–4356, 2002.
- [182] Stephen J Pearton, Richard Deist, Fan Ren, Lu Liu, Alexander Y Polyakov, and Jihyun Kim. Review of radiation damage in gan-based materials and devices. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, 31(5):050801, 2013.
- [183] SA Vitusevich, N Klein, AE Belyaev, SV Danylyuk, MV Petrychuk, RV Konakova, AM Kurakin, AE Rengevich, A Yu Avksentyev, BA Danilchenko, et al. Effects of γ -irradiation on algan/gan-based hems. *physica status solidi (a)*, 195(1):101–105, 2003.
- [184] Hiroshi Yasuda, Gautam D Badhwar, Tatsuto Komiyama, and Kazunobu Fujitaka. Effective dose equivalent on the ninth shuttle–mir mission (sts-91). *Radiation research*, 154(6):705–713, 2000.