

12-31-2010

A Framework for Determining the Reliability of Nanoscale Metallic Oxide Semiconductor (MOS) Devices

Wilkistar Otieno
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [American Studies Commons](#), [Industrial Engineering Commons](#), [Other Environmental Sciences Commons](#), and the [Sustainability Commons](#)

Scholar Commons Citation

Otieno, Wilkistar, "A Framework for Determining the Reliability of Nanoscale Metallic Oxide Semiconductor (MOS) Devices" (2010). *USF Tampa Graduate Theses and Dissertations*.
<https://digitalcommons.usf.edu/etd/3499>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

A Framework for Determining the Reliability of Nanoscale Metallic Oxide
Semiconductor (MOS) Devices

by

Wilkistar Otieno

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: O. Geoffrey Okogbaa, Ph.D.
Tapas K. Das, Ph.D.
Chris P. Tsokos, Ph.D.
Jing Wang, Ph.D.
Ashok Kumar, Ph.D.
Shekhar Bhansali, Ph.D.
Susana Lai-Yuen, Ph.D.
Sanjukta Bhanja, Ph.D.

Date of Approval:
August 11, 2010

Keywords: Nanoreliability, Dielectric, Accelerated Degradation, Kernel Density
Estimates, Bayesian Density Estimates

Copyright©2010, Wilkistar Otieno

DEDICATION

To my parents for ensuring that we all got educated in the fields of our choice.

To my mother Elsa Atieno Otieno, for being my first tutor and my father Samson Odongo Otieno for instilling in me the love for knowledge. To my brothers Fred and Guka, and sister Nelly for their incredible support throughout this long journey.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	v
ABSTRACT	ix
PREFACE	xi
CHAPTER 1 INTRODUCTION	1
1.1 Nanoreliability	1
1.2 Transistor Operation	4
1.3 Scaling Trend in Transistor Technology	6
1.3.1 Effect of SiO_2 -Based Transistor Scaling and its Limitations	8
1.3.2 SiO_2 Gate Scaling Challenges	11
1.4 Why High-k Dielectric Films?	12
1.4.1 Propositions for Extending MOSFET Technology: Alternative Gate Dielectrics	12
1.4.2 Performance Related Challenges in the Introduction of High-k Dielectric Material	16
CHAPTER 2 DIELECTRIC PERFORMANCE, FAILURE (BREAKDOWN) AND TESTS	18
2.1 Dielectric Performance and Electrical Characterization	18
2.1.1 Gate Dielectric Breakdown	19
2.1.2 Soft versus Hard Breakdown	20
2.2 Dielectric Failure Tests	22
2.2.1 Accelerated Degradation Tests	24
CHAPTER 3 RESEARCH MOTIVATION AND RESEARCH QUESTIONS	28
3.1 Why Nanoreliability?	28
3.2 Nanoreliability Framework	30
3.3 Statement of the Problem and Research Questions	31

CHAPTER 4	SIMULATION OF DIELECTRIC BREAKDOWN	34
4.1	Modes of Dielectric Breakdown	34
4.2	Current Leakage Due to Quantum Tunneling	34
4.3	Stress-Induced Defect Generation: Physical Models	35
4.3.1	Electron-Energy Dissipation Model (1/E model)	36
4.3.2	The Thermo-Chemical Model (E-model)	38
4.3.3	Defects Identification and Quantification	40
4.4	Defect-Based 3D Breakdown Simulation Model	41
4.4.1	3D Breakdown Simulation Model Details	42
4.4.1.1	3D Simulation Model Assumptions	46
4.4.1.2	3D Simulation Model Pseudocode	47
4.5	Simulation Results	51
4.5.1	3D Simulation Model Extension to the Breakdown Time of the Dielectric	53
4.6	Conclusion	54
CHAPTER 5	NONPARAMETRIC ESTIMATION OF FAILURE TIME DISTRIBUTION: USING KERNEL DENSITY METHODOLOGY	58
5.1	Background	58
5.2	Choice of Kernel Estimate: Kernel Fitness Measure	62
5.3	Determination of Mean Square Error (MSE) of a Kernel Density Estimator	63
5.3.1	Asymptotic MSE Approximation	64
5.4	Bandwidth Selection	69
5.5	Variable Optimal Bandwidth Selection	70
5.5.1	Optimum Bandwidth Selection Methodology	72
5.5.1.1	Constant Bin Width Selection Procedure	73
5.5.1.2	Variable Bandwidth Selection Procedure	74
5.6	Numerical Results	75
5.7	Reliability Inference and Numerical Results	76
5.8	Conclusion	77
CHAPTER 6	BAYESIAN INFERENCE	79
6.1	Background	79
6.2	Choice of Likelihood Function	80
6.3	Two-Parameter Weibull Distribution in the Context of Dielectric Breakdown	82
6.3.1	Weibull Shape Parameter	82
6.3.2	Weibull Scale Parameter	83
6.3.2.1	Arrhenius-Weibull Model	83
6.3.3	Weibull Plot	85
6.4	Choice of Prior Distributions	89
6.5	Hierarchical Bayesian Implementation	92

6.5.1	Limits of the Parameters and Hyperparameters	96
6.6	Posterior Computation	97
6.6.1	MCMC Simulation	98
6.7	MCMC Simulation Results	103
6.7.1	Comparison of Hierarchical Bayesian Model I and II	103
6.8	Sensitivity Analysis	106
6.9	Reliability Inference	113
6.9.1	Weibull Maximum Likelihood Estimates (MLE)	113
6.9.2	Weibull Least Square Error (LSE) Estimates	113
6.9.3	Extrapolating Dielectric Characteristic Life from Accelerated Level to Normal Use Condition	116
6.9.4	Mean Time to Failure (MTTF) Extrapolation	117
CHAPTER 7	RESULTS AND MAJOR CONTRIBUTIONS	121
7.1	3D Simulation Model	121
7.2	Kernel Density Estimation	123
7.3	Bayesian Inference	123
7.4	Contributions	125
CHAPTER 8	SUGGESTIONS FOR FUTURE RESEARCH	128
8.1	Dielectric Failure Simulation	128
8.2	Nonparametric Density Estimation	129
8.3	Bayesian Reliability Inference	129
REFERENCES		131
APPENDICES		140
Appendix A	Graphical Results: Nonparametric Normal Kernel Probability Density Estimates	141
ABOUT THE AUTHOR		End Page

LIST OF TABLES

Table 1.1	Progress of Intel's processor features	7
Table 1.2	Summary of dielectric materials [1]	14
Table 3.1	Potential percentage value of nanotechnology per sector in 2015	29
Table 5.1	Relative inefficiencies of common kernels	68
Table 5.2	KDE constant and variable bandwidth for failure data at different stress levels	75
Table 6.1	Log-likelihood values for different distributions at given stress levels	80
Table 6.2	Results of K-S goodness-of-fit tests for Weibull, log-normal and gamma distributions	81
Table 6.3	Shape parameter estimates of simulated failure time data	83
Table 6.4	Posterior results:Bayesian hierarchical model I	104
Table 6.5	Posterior results:Bayesian hierarchical model II	105
Table 6.6	Table of priors for sensitivity analysis	106
Table 6.7	Sensitivity analysis	109
Table 6.8	Dielectric characteristics life estimates (seconds)	115
Table 6.9	Weibull shape parameter estimates	116
Table 6.10	Comparison of MTTF estimates	118
Table 6.11	Extrapolated MTTF 95% confidence interval at 6 MV/cm	118
Table 6.12	Extrapolated MTTF 95% confidence interval at 2 MV/cm	119

LIST OF FIGURES

Figure 1.1	Transistor classification	4
Figure 1.2	nMOSFET in the OFF-state (fig a) and ON-state (fig b)	5
Figure 1.3	Moore's law: Intel micro-processor realization and projection	6
Figure 1.4	Trend of the average selling price of per bit of DRAM [2]	7
Figure 1.5	Clock (switching) speeds of microprocessors with time [2]	8
Figure 1.6	A schematic representation of a MOS structure	10
Figure 1.7	Charge distribution in the MOS dielectric	10
Figure 2.1	Schematic representation of a MOSFET	19
Figure 2.2	A schematic representation of the occurrence of a SBD that eventually becomes a HBD path during the breakdown process [3]	21
Figure 2.3	Schematic representation of a High-k MOS Capacitor (MOSCAP) structure for mercury probe measurements	25
Figure 3.1	Development of nanotechnology markets worldwide by regions (NSF 2015 forecast)	29
Figure 3.2	A general reliability framework	31
Figure 4.1	A schematic energy band diagram of the MOS structure where Φ is the energy barrier height at the dielectric/substrate interface, V_{ox} is the potential drop across dielectric film and V_G is the applied gate voltage [1]	35
Figure 4.2	Generation of a Pb trapped charge defect after impact ionization [1]	37
Figure 4.3	Generation of a Pb charged defect by the thermal-chemical process [1]	38

Figure 4.4	Division of the dielectric in a lattice of 3D cells with lattice constant of a_0^3	43
Figure 4.5	Point defect generation forming a conduction path	43
Figure 4.6	Point defects are insufficient to form a conduction path	43
Figure 4.7	3D representation showing some defective cells	46
Figure 4.8	3D simulation size	48
Figure 4.9	2D representation showing defect filled cells as 1s, and defect-free cells as 0s	48
Figure 4.10	2D representation showing clusters of communicating cells	49
Figure 4.11	2D representation showing an insulating array	49
Figure 4.12	2D representation showing a failed (conducting) array	50
Figure 4.13	Critical defects densities for 1, 2, 3, 4 & 5 nm dielectric thickness	52
Figure 4.14	Effect of the critical number of paths on critical defect density	53
Figure 4.15	Simulation failure times at 2.3 V	55
Figure 4.16	Simulation failure times at 3.8 V	56
Figure 4.17	Simulation failure times at 5V	56
Figure 5.1	Procedure for bandwidth optimization	73
Figure 5.2	KDE with constant (right) and variable (left) bandwidth for failure data at 8.1 MV/cm electric field stress level	76
Figure 5.3	Reliability and cdf estimates at 8.1 MV/cm electric field stress level	77
Figure 6.1	Weibull plot of simulated failure data for 1, 2, 3, 4 & 5 nm dielectric thickness with 26 degrees of communication	87
Figure 6.2	Weibull plot of simulated failure data for 1, 2, 3, 4 & 5 nm dielectric thickness with increased defect generation	88
Figure 6.3	Weibull plot of actual dielectric failure data from accelerated degradation tests	89

Figure 6.4	Graphical representation of the 3-stage hierarchical Bayesian framework	93
Figure 6.5	Trace plots indicating convergence	101
Figure 6.6	Schematic presentation of model I hierarchical Bayesian model	102
Figure 6.7	Schematic presentation of model II hierarchical Bayesian model	102
Figure 6.8	$\ln(\theta)$ posterior kernel density	108
Figure 6.9	β posterior kernel density	110
Figure 6.10	α posterior kernel density	111
Figure 6.11	γ posterior kernel density	112
Figure 6.12	Graph of extrapolated MTTF	119
Figure A.1	Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 8.1 MV/cm electric field stress level	141
Figure A.2	Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.9 MV/cm electric field stress level	142
Figure A.3	Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level	142
Figure A.4	Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level	143
Figure A.5	Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level	143
Figure A.6	Reliability (left) and cdf (right) estimates at 8.1 MV/cm electric field stress level	144
Figure A.7	Reliability (left) and cdf (right) estimates at 7.9 MV/cm electric field stress level	145
Figure A.8	Reliability (left) and cdf (right) estimates at 7.7 MV/cm electric field stress level	145

Figure A.9	Reliability (left) and cdf (right) estimates at 7.5 MV/cm electric field stress level	146
Figure A.10	Reliability (left) and cdf (right) estimates at 7.3 MV/cm electric field stress level	146
Figure A.11	Reliability (left) and cdf (right) estimates at 7.1 MV/cm electric field stress level	147

A Framework for Determining the Reliability of Nanoscale Metallic Oxide Semiconductor (MOS) Devices

Wilkistar Otieno

ABSTRACT

An increase in worldwide investments during the past several decades has propelled scientific breakthroughs in nanoscience and technology research to new and exciting levels. To ensure that these discoveries lead to commercially viable products, it is important to address some of the fundamental engineering and scientific challenges related to nanodevices. Due to the centrality of reliability to product integrity, nanoreliability requires critical analysis and understanding to ensure long-term sustainability of nanodevices and systems. In this study, we construct a reliability framework for nanoscale dielectric films used in Metallic Oxide Semiconductor (MOS) devices. The successful fabrication and incorporation of metallic oxides in MOS devices was a major milestone in the electronics industry. However, with the progressive scaling of transistors, the dielectric dimension has progressively decreased to about 2nm. This reduction has had severe reliability implications and challenges including: short channeling effects and leakage currents due to quantum-mechanical tunneling which leads to increased power dissipation and eventually temperature related gate degradation.

We develop a framework to characterize and model reliability of recently developed gate dielectrics of Si-MOS devices. We accomplish this through the following

research steps: (i) the identification of the failure mechanisms of Si-based high-k gates (stress, material, environmental), (ii) developing a 3-D failure simulation as a way to acquire simulated failure data, (iii) the identification of the dielectric failure probability structure using both kernel estimation and nonparametric Bayesian schemes so as to establish the life profile of high-k gate dielectric. The goal is to eventually develop the appropriate failure extrapolation model to relate the reliability at the test conditions to the reliability at normal use conditions.

This study provides modeling and analytical clarity regarding the inherent failure characteristics and hence the reliability of metal/high-k gate stacks of Si-based substrates. In addition, this research will assist manufacturers to optimally characterize, predict and manage the reliability of metal high-k gate substrates. The proposed reliability framework could be extended to other thin film devices and eventually to other nanomaterials and devices.

PREFACE

My deepest appreciation goes to my major professor Prof. O. Geoffrey Okogbaa for his help and encouragement during my graduate work. I specifically thank him for nurturing me into a meticulous researcher, who is able to develop, structure and present ideas in an understandable way. I also wish to express my sincere gratitude to all my dissertation committee members: Prof. Chris Tsokos, Prof. Ashok Kumar, Prof. Shekhar Bhansali, Prof. Sanjukta Bhanja, Prof. Susana Lai-Yuen and Prof. Jing Wang for their valuable feedback.

I'm deeply indebted to Prof. Tapas Das who has been a mentor and together with my major professor, provided funds that enabled me complete my graduate studies, through the GK-12 STARS (NSF-DGE-0139348 and 0638709) program at USF's College of Engineering. I am also very grateful to Prof. Jose Zayas-Castro, the USF-Industrial and Management Systems Engineering (IMSE) Department Chair for giving me the opportunity to serve in the Department and more so, to teach Probability and Statistics for Engineers. I am grateful for the support I received from Ms. Gloria Hanshaw and Ms. Jackie Stephens, both of whom are staff in the IMSE department.

In the course of this research, we corresponded with Prof. Jordi Sune of the University of Barcelona, Spain, and Dr. John Suehle and Dr. Kin Cheung, both of whom are researchers at the National Institute of Standards and Technology (NIST). I thank them all for their insight.

My journey through graduate school would not have been any easier were it not for the supportive friends I made during my doctoral program. I am so grateful to my Ugandan sister Sarah Namulondo, for providing a beacon of hope, and most of all, for teaching me the art of writing. I also thank Sorina Susnea, Leyla Goldsmith, Selasi Blavo, Vishnu Nanduri, Chaitra Gopalappa, Diana Prieto, Patricio Rocha and all the INFORMS-USF Chapter members for the fun (international potlucks) and encouragement they provided. Special thanks go to Tom Owiti; while you may not have understood the pressure I was facing at times, your constant love and support kept me going. To my family: mum, dad, Fred, Alison, Guka, Nelly, Robert, Verna, Regina and Chepte - your love and support has meant a lot to me. Lastly and most importantly, I thank the Almighty for being my rock and refuge when all other ground around me appeared to crumble.

CHAPTER 1

INTRODUCTION

1.1 Nanoreliability

Nanoreliability is the probabilistic measure of the ability of a nanoscale material, device or systems to maintain its intended functionality at predetermined use environmental conditions and time. The physical, chemical, thermal, electrical and biological properties of materials differ significantly at the nanoscale level. Thus such properties should be considered in characterizing failure and reliability. Today, nanostructured materials and devices serve critical roles in newly developed products. As a consequence, there are unprecedented levels of integration of emerging nanomaterials, devices and technologies, all of which pose severe challenges for the testing, reliability, and metrology requirements needed to support such development [4].

Advances in reliability research and development have, in the past, enhanced product quality and performance. The gains in this context occurred largely through the identification of the failure mechanisms and the translation of this information into remedies that impact design, manufacturing, and intervention regimens. With the paradigm shift from conventional material and manufacturing processes to nanomaterials and nanomanufacturing, it is imperative that new tools and techniques be implemented that are specifically aligned with the realities of nanodevices. Reliability analysis typically involves testing, measurement, modeling, and prediction of the various aspects of a product's life cycle with the ultimate goal of establishing

survivability and performability requirements. In the context of nano scale, there are several challenges that confront reliability researchers more so than traditional systems, such as the identification of the underlying physics of failure mechanisms, the development of appropriate testing methods and instrumentation at the nanoscale level to acquire failure data, characterization of material and device failure, and the development of reliability models that incorporate the physics of failure, to enhance operability requirements. The major thrust of nanoreliability is anchored around several overarching themes, such as massive parallelism of components to overcome device failures, performability, and the life cycle of those devices.

The ultimate goal of nanoreliability is to determine the availability of nanodevices and nanomaterials under normal use conditions. Elsayed [5] identifies three reliability inference procedures (and models), namely: (i) physics-experiment-based models, (ii) physics-statistics-based models, and (iii) statistics-based models . In this study, we incorporate these model categories into the reliability framework of CMOS high-k thin dielectric films as follows:

1. The physics-experimental models involve the use of failure test experiments to estimate failure times. These tests involve applying different stresses that affect the performance of the nanomaterial or device. Traditionally, failure data is obtained by assessing the performance of a product throughout its life. Life data analysis involves the use of times-to-failure data obtained under normal operating conditions to quantify the life profile of the product or device. However, in many instances such life data is very difficult, if not impossible to obtain especially when the products are designed to operate without failure for several years. Given this difficulty, methods and approaches have been developed to accelerate device failure. Such accelerated failure tests (ALTs) involve over stressing the devices by exposing them to severe conditions than those experi-

enced under normal conditions. We discuss the concept of accelerated testing in chapter 2 of this dissertation.

2. The physics-statistics-based models relate known chemistry and physics of failure principles and the ensuing failure rate. These models are used to theoretically estimate failure rates and therefore failure times. In this study, the failure-physics and failure-rate relationships of high-k thin films are presented in chapter 2. We use this relationship to develop a theoretical 3-dimensional failure simulation model for high-k dielectric thin films in chapter 4.
3. The statistics-based models are generally used when the relationship between failure times and the applied stresses is unknown, and difficult to derive from the known physics or chemistry-based failure principles. In this case, the devices are tested, and the resulting failure times are used to estimate the underlying probability structure that describes the life characteristics of the device such as the failure density function, failure rate, the cumulative failure and the reliability functions. Both parametric and nonparametric statistical procedures are used to estimate these life characteristic measures. In this research, we specifically apply nonparametric and parametric methodologies to make reliability inference as detailed in chapters 5 and 6 respectively.

Overall, the accuracy of the inference procedure applied has profound effects on the reliability estimates. In this dissertation we develop a framework to characterize and model the reliability of dielectric thin films, and validate the models in the framework using real data obtained from Luo et al. [6].

1.2 Transistor Operation

Transistors are semiconductor devices that are used in most electronic devices, and mainly function as signal amplifiers or as solid-state-switching components. There are three main chronological classifications of transistors as shown in figure 1.1. They include: (i) Bipolar Transistors (BTs) which are considered current driven and have a relatively low input impedance, (ii) Field effect Transistors (FETs), also known as voltage driven devices and have a high input impedance, and (iii) Insulated Gate Transistors Insulated Effect Transistors (IGBTs), which is the newest and is a hybrid combining the BTs and the FETs. FETs are further classified into: (a) Junction Field Effect Transistors and (b) Metallic Oxide Semiconductor Field Effect Transistors (MOSFET) [7].

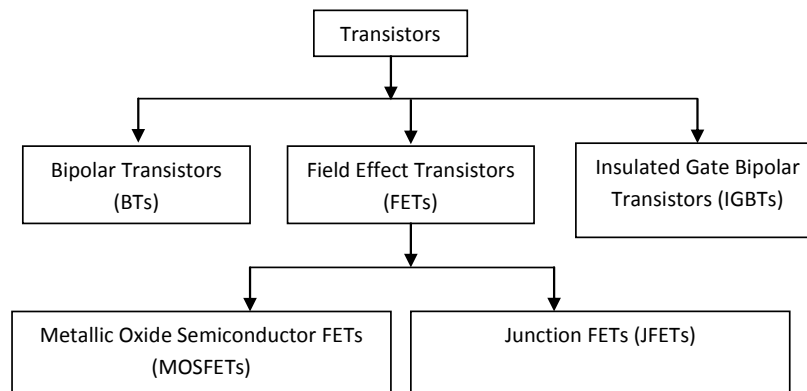


Figure 1.1. Transistor classification

Figures 1.2 a and b are cross sectional views of an n-channel MOSFET, which has highly doped n+ source and drain terminals. The gate terminal, the drain and the source interconnections are metallic or made of highly doped polysilicon. The insulator is generally made of a dielectric material, traditionally SiO_2 , and the substrate is made of a lightly p doped semiconductor.

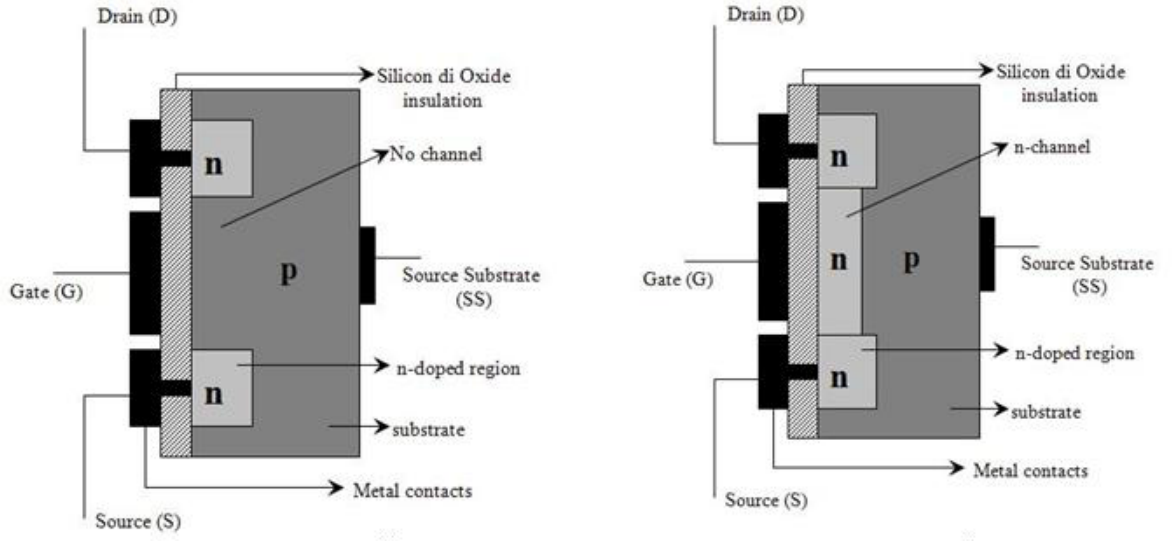


Figure 1.2. nMOSFET in the OFF-state (fig a) and ON-state (fig b)

The operation of the MOSFET is induced by modulating the charge concentration in the MOS capacitor (the area defined by the gate terminal, the dielectric film and the substrate). For the nMOSFET 1.2 a, when the applied gate voltage (V_G) is positive and higher than the threshold voltage V_{TH} , an electric field is induced over the insulator which attracts negative charges (electrons) at the insulator/substrate interface. The electron concentration increases with an increase in the gate bias until the electrons form an inversion conducting channel between the source and the drain (see figure 1.2 b). When the bias is reduced, the inversion layer is cut-off, and a small amount sub threshold current may leak through from the source to the drain.

The opposite process occurs in the case of a pMOSFET, in that, when V_G is less than V_{TH} , holes are attracted to the insulator-substrate interface. When the bias is reduced further, the hole concentration increases in the interface therefore forming a conduction channel. Shrinkage of transistor dimensions has had several implications in the structure and functionality of the transistor within semiconductor devices, some of which will be discussed in the subsequent sections.

1.3 Scaling Trend in Transistor Technology

In the paper, *Cramming more Components onto Integrated Circuits (IC)*, Moore noted that the objective of miniaturization is to include as many electronic components as possible in the smallest space possible and with the least weight implications as possible [8]. He postulated that the number of transistors on an IC would double every two years [8]. As shown in figure 1.3 transistor trend in recent years, as recorded by Intel, support Moore's prediction, as we have progressed from 10000 transistors on a chip in 1975 to over 1 billion transistors on the same chip size in 2008.

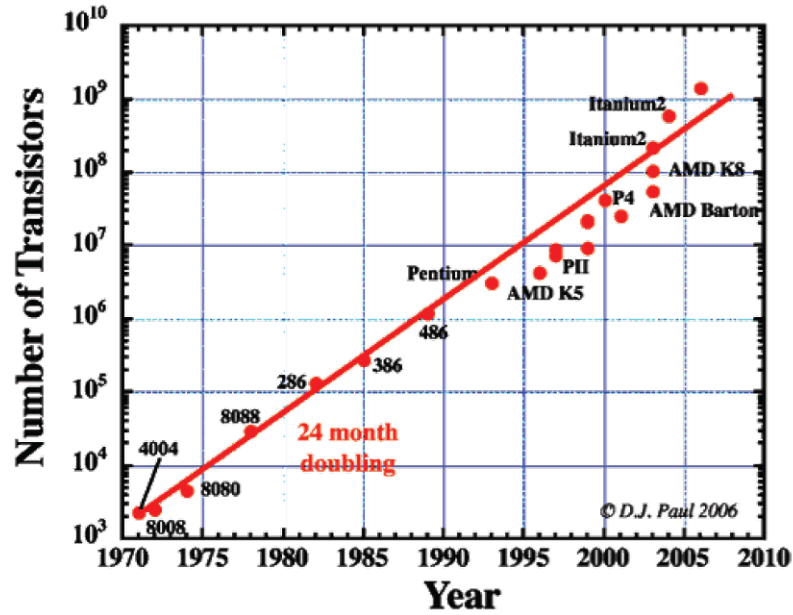


Figure 1.3. Moore's law: Intel micro-processor realization and projection

Moore's predictions have been made possible by miniaturization of transistors used in ICs. Today semiconductor industries can manufacture logic devices that incorporate up to 40 million transistors into a single circuit [9]. According to the Intel Developer Forum [9], Intel has had a progressive reduction in the size of computer processors as shown in table 1.1, in the process recording thousands of millions of electronic components on a chip-also known as Very Large Scale Integration (VLSI).

Table 1.1. Progress of Intel's processor features

1 st Production	1997	1999	2001	2003	2005	2007
Process name	P858	P858	Px60	P1262	P1264	P1266
Processor name	Pentium II	Pentium III Xeon	Pentium 4	Pentium M	Pentium D	Pentium D-2
Lithography	0.25 μ m	0.18 μ m	0.13 μ m	90nm	65nm	45nm
Gate thickness	0.20 μ m	0.13 μ m	<70nm	<50nm	<35nm	<25nm
Wafer diameter (mm)	200	200	200/300	300	300	300
Number of Transistors (Million)	9	50	70	330	700	1000

It is evident from table 1.1 that the electronic industry has embraced the era of nanotechnology and is manufacturing millions of electronic components at a phenomenal scale [9]. The key productivity features that have been exhibited in the semiconductor industry include decreasing unit cost of production and the increasing clocking (switching) frequency of nano-electronic components as shown in figure 1.4 and 1.5 [2].

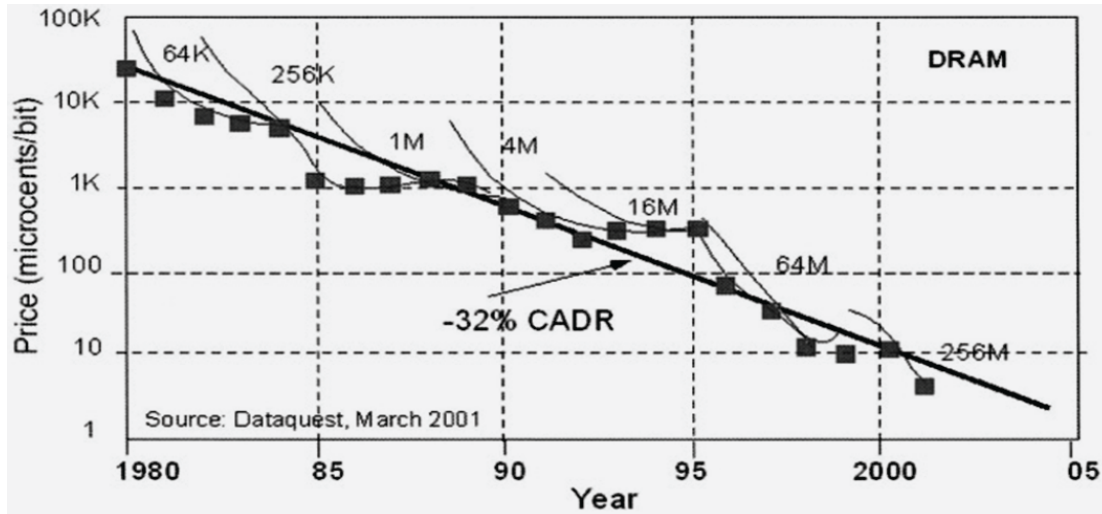


Figure 1.4. Trend of the average selling price of per bit of DRAM [2]

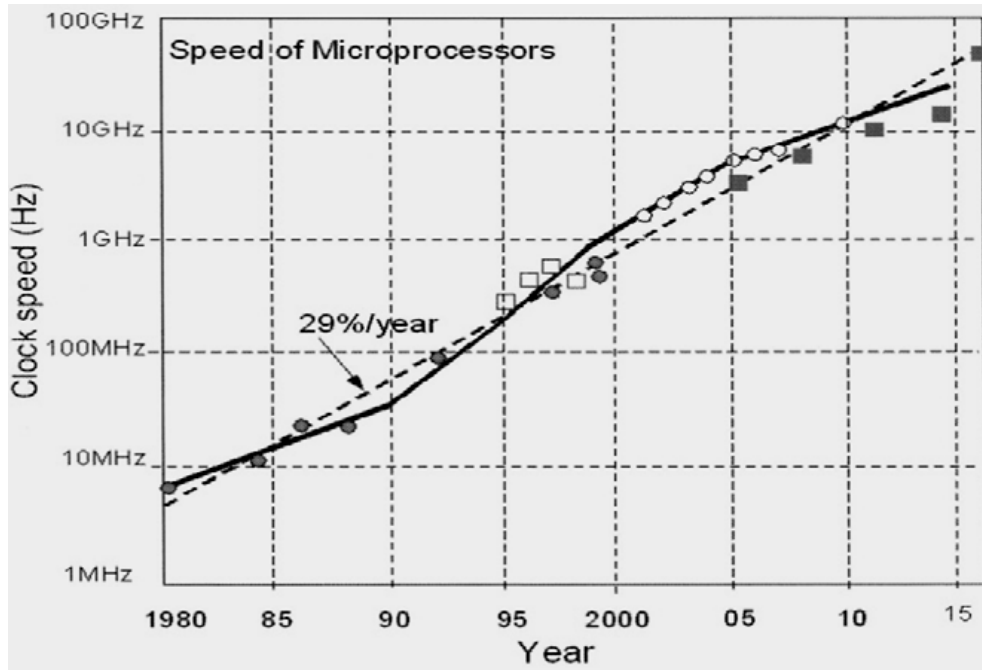


Figure 1.5. Clock (switching) speeds of microprocessors with time [2]

The scaling trend and inexpensive nano electronic components are giving rise to relatively inexpensive yet extraordinarily powerful integrated circuits and electronic devices [10], and this has been the driving force behind the successes in the electronic industry.

1.3.1 Effect of SiO_2 -Based Transistor Scaling and its Limitations

The increase in the number of transistors on a chip has been made possible by the progressive reduction (scaling) of transistor dimensions among which is the dielectric thickness. The successful fabrication and incorporation of metallic oxides in transistors was a major milestone and significant achievement in the electronics industry, and the use of Silicon dioxide as the choice dielectric certainly encouraged continuous scaling. Amorphous SiO_2 can easily be thermally grown on Si substrate (wafers) with precise control of its thickness [1]. SiO_2 naturally forms a stable oxide

interface with the Si substrate and this results in very low intrinsic interfacial defect density. Houssa [1] indicates that even the few defects that may form are efficiently passivated through post-metalization annealing processes (see Deen et al. [11] for a detailed study of post-deposition annealing processes). In addition, SiO_2 forms an interface with a large barrier height of approximately 3 eV and it also has an inherent large band-gap, approximately 9 eV in magnitude, which offers excellent electrical isolation properties [12]. A large band-gap and wide barrier height means that the energy band offsets between the conduction and the valence bands is large, therefore the breakdown field is also large (to the order of 13 MV/cm). Such a high breakdown field is indicative of the amount of input voltage that the dielectric can withstand before breakdown [1]. Due to its superior electrical and mechanical properties SiO_2 was used as the gate dielectric in MOSFET fabrication for decades until the thickness was scaled down to 1.5 nm, beyond which reliability challenges arose.

From an electrical point of view, the MOS structure behaves like a parallel plate capacitor, consisting of a dielectric material sandwiched between the gate and the substrate electrode. Figure 1.6 shows a schematic representation of the MOS structure. When V_G is applied to the gate electrode, charges accumulate on the metallic gate electrode as shown in figure 1.7, and are compensated for by opposite charges in the substrate interface region. In a MOS transistor, these compensating charges are responsible for forming the channel that connects the source to the drain as was discussed earlier.

The capacitance (C), of the MOS capacitor is given by:

$$C = \frac{\epsilon_r \epsilon_0 A}{t_{diel}} \quad (1.1)$$

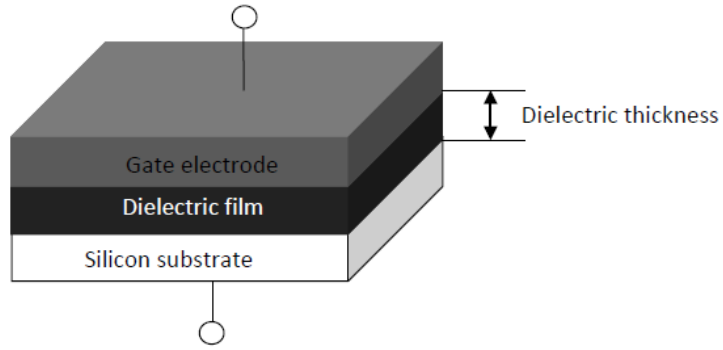


Figure 1.6. A schematic representation of a MOS structure

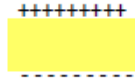


Figure 1.7. Charge distribution in the MOS dielectric

where ϵ_r is the relative static permittivity (also known as the dielectric constant), A is the active area of the capacitor, t_{diel} is the dielectric thickness, and ϵ_0 is the vacuum permittivity.

Evidently from equation 1.1, reducing the insulator (dielectric) thickness increases the capacitance of the insulator, and consequently increases the transistor drain current, thereby improving the transistor speed/performance. Within the past decade, the dielectric thickness has steadily decreased from a range of 1.3 to 1.6 nm in 2001, to a current (2010) value of 0.6 to 1.1 nm [1]. This reduction has had severe reliability implications and challenges such as short channeling effects, leakage currents either due to quantum-mechanical tunneling, or as a result of conduction paths formed by stress-induced defects within the dielectric film [13], [14] and [15].

1.3.2 SiO_2 Gate Scaling Challenges

1. Sub threshold leakage current:

One of most stringent limitations toward further miniaturization of dielectric films concerns the leakage current that flows through the MOS stack. When the dielectric film thickness is decreased typically beyond 3nm, charge carriers are able to flow through the film by a quantum mechanical tunneling mechanism. Houssa et al. [1] indicate that the tunneling probability increases exponentially with decreasing dielectric thickness. This leakage may occur during the off-state also known as the depletion stage. Leakage current during the depletion stage results in increased dissipated power, which is proportional to the product of the leakage current and the supply voltage. Increased power dissipation results in temperature associated transistor degradation [1].

2. Dielectric reliability:

Another issue related to dielectric scaling concerns reliability aspects. During MOSFET operation, an electric field is induced within the transistor channel region, which activates charge carriers to flow from the source to the drain terminals. As the carriers traverse the channel, they release some charged species which penetrate the dielectric causing defects within the dielectric bulk and interface (more details will be given in chapter 4). When a critical defect density is reached, defects align to form a conduction path through which electrons flow from the substrate to the gate electrode. This electric short leads to a dielectric breakdown phenomenon that compromises gate reliability.

3. Doping Fluctuations:

The amount of dopant atoms in the channel region reduces with transistor scaling. This reduction, coupled with the random nature of dopant distribution

within the channel results in significant variations in the sub-threshold voltage. For instance, a gate measuring 100 nm in length by 400 nm in width on average has 1000 dopant atoms in its depletion region [16]. When the gate length is reduced to 25 nm, the number of dopant atoms reduces to approximately 120, and this greatly varies the threshold voltage from one device to another. To abate this, retrograde doped channels have been developed, which consist of a low followed by a high doping profile of the semiconductor substrate [16].

1.4 Why High-k Dielectric Films?

1.4.1 Propositions for Extending MOSFET Technology: Alternative Gate Dielectrics

From equation 1.1, a reduction in dielectric thickness results in increased MOS capacitance and eventually reduced device delays. Currently, the dielectric thickness ranges between 1 to 1.5 nm and it is foreseen that by 2016, this thickness will reach a record 0.4 to 0.5 nm [1]. This leads to increase in leakage currents and hence increased dissipated power and heat within the dielectric region. SiO_2 films of about 1.5 nm in thickness exhibit approximately 1 A/cm^2 of current density at an input voltage of 1 V [1]. Since current leakage density increases exponentially with decreasing dielectric thickness, a 1.0 nm thick oxide is expected to exhibit up to 100 A/cm^2 of current density at the same input gate voltage [1]. Keeping this in mind, SiO_2 would only be scaled down to 0.8 nm for high performance applications [1]. From equation 1.1, the alternative way of increasing MOS capacitance is by increasing the dielectric constant ϵ_r , which is achievable when alternative dielectric films are used to replace SiO_2 . These alternative dielectric materials also known as high-k materials, have relatively higher permittivity ϵ_r than SiO_2 .

Even though the first field effect semiconductor devices containing zirconium (Zr) and hafnium (Hf)-based oxynitride gate dielectrics were patented in 2000, the first commercial high-k MOSFETs were successfully built by Intel corporation in 2003 [17]. Other candidate dielectric materials that have been studied include derivatives of Zr and Hf alloys with interfacial SiO_2 and Al_2O_3 films. The dielectric constant of ZrO_2 and HfO_2 are in the range of 17 – 25, compared to that of SiO_2 which varies between 3 to 5. High-k materials exhibit slightly higher conduction band energy levels of 1.4 to 1.5 eV compared to SiO_2 whose conduction band energy level is about 1 eV [1]. Lower conduction band energy levels are desired to ensure low current leakage. Intel realized that the introduction of the high-k material reduces the gate leakage by 100-fold [17].

By using these high-k dielectric materials, thicker layers of dielectric films can be used to achieve the same capacitance while at the same time reducing the leakage current, which consequently increases dielectric reliability. The equivalent dielectric thickness is defined as the thickness of SiO_2 layer that would be required to achieve the same capacitance density as the alternative high-k dielectric [1]. From equation 1.1, assuming that the capacitance C , and the cross sectional area A , are held constant, the equivalent SiO_2 thickness $EOT = t_{eq}$ is obtained as follows:

$$\frac{t_{eq}}{\epsilon_{r,SiO_2}} = \frac{t_{high-k}}{\epsilon_{r,high-k}} \quad (1.2)$$

where ϵ_{r,SiO_2} and $\epsilon_{r,high-k}$ are the relative dielectric constants of SiO_2 and high-k material respectively, and t_{eq} and t_{high-k} are the equivalent SiO_2 and high-k dielectric thickness respectively.

Table 1.2 is a summary of some of the current dielectric materials, indicating their dielectric constants, band gaps E_g eV, conduction band energy levels E_c eV, stability on silicon substrate, and their crystallographic structures.

Table 1.2. Summary of dielectric materials [1]

Material	Dielectric constant (k)	E_g (eV)	E_c (eV)	Stability on Silicon	Crystallographic structure
SiO ₂	3.9	9	3.1	Yes	Amorphous
Si ₃ N ₄	7.8	5.3	2.4	Yes	Amorphous
TiO ₂	80	3.05	0.05	No	Tetragonal
ZrSi _x O ₇	~10	~6.5	1.5	Yes	Amorphous
HfO ₂	25	~6	1.5	Yes	Monoclinic, tetragonal, Cubic
ZrO ₂	25	5.8	1.4	Yes	Monoclinic, tetragonal, Cubic
HfSi _x O ₇	~10	~6	1.5	Yes	Amorphous
Al ₂ O ₃	10	9.9	2.8	Yes	Amorphous
Al _x Zr _{1-x} O ₂	14				
Y ₂ O ₃	15	6	2.3	Yes	Cubic
La ₂ O ₃	30	6	2.3	Yes	Hexagonal
Pr ₂ O ₃	31			Yes	
Ta ₂ O ₅	26	4.4	0.3	No	Orthorhombic

When Zr or Hf-based dielectrics are deposited on the Si substrate, a thin interfacial layer of low-k SiO_2 or SiM_xO_y film is formed, (M stands for Zr or Hf). This interfacial layer, whose properties are represented by the subscript low-k, forms either during

deposition or post-deposition annealing process. As a result, the capacitance of the gate stack, C_{total} is expressed as:

$$\frac{1}{C_{total}} = \frac{1}{C_{low-k}} + \frac{1}{C_{high-k}} \quad (1.3)$$

The equivalent oxide thickness is therefore:

$$t_{eq} = \left(\frac{\epsilon_{r,SiO_2}}{\epsilon_{r,high-k}} \right) t_{low-k} + \left(\frac{\epsilon_{r,SiO_2}}{\epsilon_{r,high-k}} \right) \quad (1.4)$$

This additional interfacial film in effect increases the actual dielectric layer. For instance, from equation 1.4 a gate stack formed by 0.7 nm of SiO_2 interfacial layer and 5.1 nm of ZrO_2 layer results in an equivalent oxide thickness (EOT) of 1.7 nm, as opposed to 1 nm EOT without the interfacial layer. The slight increase in the EOT due to the interfacial layer is therefore not desired, and several research publications have addressed technologies that reduce the formation of the low-k interfacial layer. Most of these research papers have considered varied passivation processes [18], [19], [20] and [21].

Generally, any suitable alternative high-k material should fulfill the following requirements [1]:

1. Good thermal stability in contact with Si substrate to prevent the formation of an unstable interfacial oxide and silicide layers.
2. Low density of intrinsic defects at the Si/dielectric interface and within the dielectric bulk, to ensure high carrier mobility in the conduction channel and low instances of dielectric breakdown.

3. Sufficiently large energy band gap to provide high energy barriers at the substrate/dielectric interface and gate(metal)/dielectric interface, thus reducing leakage currents due to electron tunneling.
4. Material compatibility with other CMOS processing and packaging thermal requirements.

1.4.2 Performance Related Challenges in the Introduction of High-k Dielectric Material

High-k materials, which are not readily compatible with silicon substrates form undesirable defects within the dielectric/Si interface. During the deposition process, high-k monolayer nucleation on the H-terminated substrate (after the last HF-dipping during deposition), is often inhibited, resulting in non-uniform and discontinuous films which create sites for trapping charges. In addition to process induced defect, trapped charges are usually generated as a result of applied stresses. For instance, in negative biased pMOSFETs, when the applied gate voltage $V_G < 0V$, charges are trapped within the dielectric bulk and at the interface. This phenomenon is known as Negative Bias Temperature Instability (NBTI) [22]. These defect sites increase threshold voltage, reduce channel carrier mobility, and induce parasitic capacitance at the interface leading to overall performance degradation. NBTI currently is among the leading causes for reduced MOSFET performance and therefore poses challenges in Very Large Scale Integrated (VLSI) circuit applications, and more specifically, high temperature applications.

Another mechanism through which defects are introduced in the Si/high-k interface is known as hot carrier injection (HCI). Damage caused by HCI results from carrier heating in the high electric field region near the drain side of the MOSFET. This

heating results in impact ionization and subsequent dielectric degradation. Similar to NBTI, HCI traps shift device metrics such as the threshold voltage, and therefore reduces overall device performance. Historically, HCI is mostly significant in nMOS-FET because electrons, which are the charge carriers on n-doped semiconductors, exhibit higher mobility due to lower effective mass compared to holes in p-doped semiconductors.

In this work, we are concerned with the defect generation phenomenon, and its implication on the performance of the gate dielectric film. In chapter 2, we discuss the background of dielectric performance and breakdown classifications. In chapter 3, we present the research questions and motivation for studying dielectric performance, and the need to develop a reliability framework for nanomaterials, using the dielectric films as an example. We will discuss the physics behind dielectric failure, and develop the 3-dimensional model to simulate dielectric failure in chapter 4. Chapters 5 and 6, consist of nonparametric and parametric approaches for dielectric reliability inference. The results of this study are summarized in chapter 7 followed by suggestions for future research in chapter 8.

CHAPTER 2

DIELECTRIC PERFORMANCE, FAILURE (BREAKDOWN) AND TESTS

2.1 Dielectric Performance and Electrical Characterization

Dielectrics are insulating materials by definition [12]. However, charge carriers are still able to tunnel through a dielectric media's conduction band. The mobility of charge carriers is approximately of the same order as that in semiconductors, so that sufficient current can be conducted if enough charge carriers are injected into the dielectric media [12]. Generally, due to the difficulty of populating charge carriers within their very low conduction bands, the resulting current is usually very low in magnitude [1]. Therefore any phenomenon that might lead to increased carrier injection into the dielectric film affects its performance.

The performance of a gate dielectric is defined as its ability to act sufficiently as an insulator between the source and the drain at specified operating conditions [6]. Typically, the reliability specification for a gate dielectric is described as; less than 0.01% of gates of size 0.1 cm^2 are allowed to fail in 10 years under 1 MV/cm electric fields and 1 V_G gate voltage [6], [23]. Figure 2.1 is a schematic presentation of the cross section of a MOSFET device.

Dielectric performance highly depends on wafer quality, process integrity and application conditions. As such, causes of dielectric failure are broadly categorized into two classes [24]: (i) intrinsic failure mechanisms, are material related, and they are associated with stress-induced defects and dislocations within the dielectric bulk and

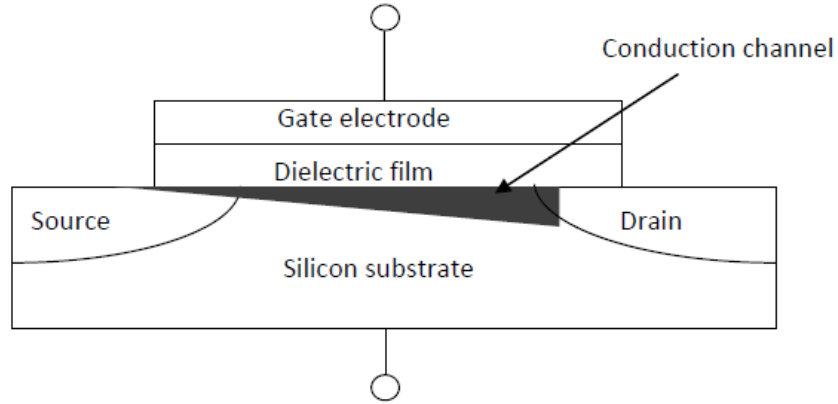


Figure 2.1. Schematic representation of a MOSFET

in the dielectric/substrate interface [24]. (ii) Extrinsic failure mechanisms are entirely process-induced, and due to defects within the dielectric bulk and in the interface that are formed during the deposition, post deposition and packaging processes [24]. Some failure mechanisms are dominant in certain operations and environments while others are prevalent across a broader range of environments. For instance, some failures may be prevalent during certain test regimens and less prevalent during normal use conditions [24]. Overall, experts agree that stress-induced intrinsic failures are typically dominant [3], [14], [25], [26] and [27].

2.1.1 Gate Dielectric Breakdown

When voltage is applied to a capacitor, the dielectric goes through several degradation processes including charge carrier injection and a multistage dielectric breakdown as a result of stress-induced defect generation [13], [28]. Sune et al. [3] define gate breakdown as the increase in gate leakage current beyond the allowable amount. Typically, the maximum tolerable standby leakage is about 1 A/cm^2 for a power supply of 1 V in CMOS logic chips, which is also the maximum tolerable leakage current for most logic applications [14]. The most common carrier injection mechanism is

quantum mechanical tunneling through energy barriers formed by the discontinuity between the substrate and the dielectric conduction bands. Unlike other injection processes such as hot carrier injection or avalanche injection, tunneling does not involve carrier injection over the energy barrier. For this reason, carrier tunneling through the barrier is considered negligible, especially for dielectric thickness beyond 7 nm. See Wu et al. [12] for more information on carrier tunneling.

The multistage defect-driven breakdown follows the following sequence: (i) initially, defects are generated at a rate that gradually increases with time, (ii) when enough defects align to form a conduction path that bridges dielectric thickness, the dielectric breaks down and the current leaks through it, causing an abrupt increase in the gate current, and (iii) whether the dielectric becomes completely conductive or remains quasi-insulating depends on the surrounding stress levels (electric field) [23] [29]. The leakage current leads to increased stand-by power dissipation within the dielectric thereby decreasing gate performance. Dielectric breakdown effects are particularly of concern in memory cells where gate leakage can result in the complete loss of stored data [28]. The critical density of defects that is enough to trigger breakdown varies with dielectric material and thickness. In this study, we will use a dielectric failure simulation model to address the relationship between the critical defect density and dielectric thickness.

2.1.2 Soft versus Hard Breakdown

Dielectric failure is a time and stress dependent degenerative stochastic event that can either be soft (soft breakdown) or hard (hard breakdown). Soft breakdown (SBD) is defined as current leakage that does not exceed the predetermined level, and therefore causes slight damage to the dielectric. SBD is usually reversed by a reduction in the electric field across the gate. This non-destructive breakdown effect is

manifested in both thin and thick dielectric films. On the other hand, hard breakdown (HBD) is destructive, irreversible and results in a sudden loss of dielectric insulation. These two modes of failure are used as indicators for dielectric failure. The conditions favoring each mode depend on the dielectric thickness, the magnitude of the stored energy in the capacitor, system impedance, and the extent of local damage caused by defect generation [24].

From a statistical point of view, these two modes of failure pose a challenge in the analysis of dielectric failure data. These challenges include: (i) whether to treat the modes of failure as separate competing events resulting in two separate failure distributions, (ii) assume that they are equally distributed and therefore merge the data irrespective of the failure mode, and (iii) whether to think of the SBD as an initial (reversible) stage of dielectric failure which eventually leads to a sudden (irreversible) HBD with time [14]. Sune et al. [3] found a solution to these challenges by studying the degradation paths followed by samples of dielectric films. Their findings are represented graphically in figure 2.2.

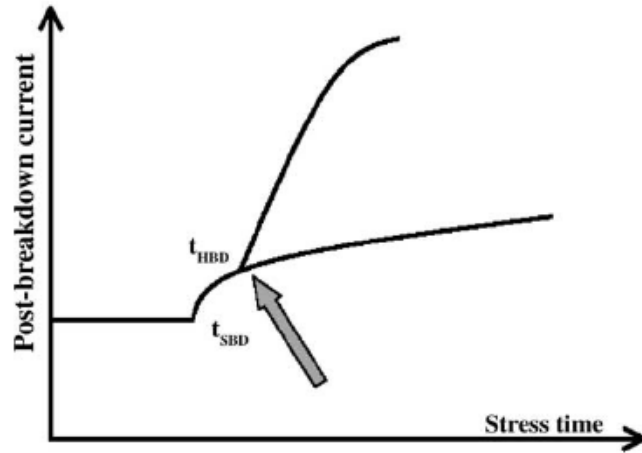


Figure 2.2. A schematic representation of the occurrence of a SBD that eventually becomes a HBD path during the breakdown process [3]

Their results show that there is a chance that SBD might lead to HBD and the corresponding probability that a defect build-up becomes a leakage pathway and transitions from SBD to HBD depends on whether a certain energy dissipation threshold has been reached. The amount of dissipated energy also determines the SBD-HBD prevalence ratio [14].

2.2 Dielectric Failure Tests

Electrical characterization tests and techniques that are required to ensure quality and improved performance of high-k materials in MOS devices have been studied extensively [1], [6], [10], [15], [30]. The results of these tests strongly depend on the sample fabrication procedures and recipes, test apparatus, test stress levels, stress time, the choice of failure indicators, and the range and resolution of the measuring instruments [31]. In this section, we will discuss the testing procedure, failure indicators and the impact of testing at elevated stress levels. However, tests were not conducted in our research facility for this study.

Dielectric reliability is defined as the probability that the dielectric will sufficiently perform as an insulator in a MOS device for the intended length of time, and at the intended operation conditions. Assessing reliability often involves making the dielectric fail by applying irreversible destructive electrical tests [12]. These tests can be done in series by stressing one sample after another, or in parallel, by stressing several samples simultaneously [12].

Necessary failure accelerators for dielectric films include voltage or current and temperature. However, major breakdown acceleration is driven by voltage or current across the dielectric film. For this reason, electrical characterization techniques that are used as means to assess dielectric reliability through breakdown detection include the amount of voltage it takes to induce dielectric breakdown, V_{bd} , the amount of

electric field that results in a breakdown E_{bd} , and the charge build-up that triggers dielectric breakdown, Q_{bd} [6]. These three electrical characteristics are candidate failure indicators, which can be monitored in an experimental setting to determine the time to breakdown t_{bd} . Elevating the test temperature decreases the activation energy thus reducing the time to failure. Typically, commercial burn-in test temperatures range between $130^{\circ}C$ to $150^{\circ}C$, while the specified maximum operating temperatures range from $70^{\circ}C$ to $125^{\circ}C$ [12], [32].

In practice, several decisions are made to scientifically design optimal accelerated failure test plans. These decisions include: (i) the type of stress to be used for instance, voltage, current, humidity and temperature, or a combination of multiple stresses, (ii) the number of test stress levels, (iii) the number of samples to be tested at each stress level, (iv) the inspection frequency, (v) the failure indicators and (vi) the stress loading strategy [33].

Typically, the stress level applied to the samples can be constant or varying (cyclically, randomly or continuously increasing) over time. The choice of the loading strategy depends on how the device is normally loaded in service [34]. There are mainly five stress loading strategies pertaining to dielectric stress tests namely, Constant Voltage Stress (CVS) test, the Ramped Voltage Stress (RVS) test, the Constant Current Test (CCS), the Constant Current Stress (CCS), and the Ramped Current Stress (RCS) [34]. More detailed discussions about the design of accelerated failure test plans as well as a rich reference of related texts are provided in [33], [35], [36], [37], [38], [39], [40], [41] and [42].

During constant stress procedures, the samples are exposed to a constant voltage or current loads. Such constant test strategies are mostly preferred because they closely represent normal use conditions, that is, most devices are assumed to experience constant stresses in service [34]. In ramped stress loading the samples are exposed

to continuously increasing voltage and current levels until they fail. While samples take shorter times to fail in ramped tests than in constant stress loading, ramped stress levels may trigger cumulative failure modes that do not occur in service.

Detection of dielectric breakdown by monitoring the current leakage and gate voltage shifts is mostly achieved through capacitance-voltage (CV) or current-voltage (IV) characteristic curves, which can be investigated using the Mercury-Probe-Technique [1]. Houssa presents a detailed discussion of the mercury-probe technique which is represented schematically in figure 2.3 [1]. The mercury probe device is a test structure that replicates the behavior of the dielectric in an actual MOS device. Instead of fabricating detailed MOS devices such as transistors, the mercury-probe technique offers a simple, quicker method for testing the dielectric films in a Metallic-Oxide-Semiconductor structure [43]. In figure 2.3 the MOS structure is made of a mercury gate electrode, a high-k dielectric film, a silicon (Si) or germanium (Ge) semiconductor substrate and a back ohmic contact which can be made of aluminum. To ensure characterization results that are devoid of fabrication errors, prior research is required in deciding optimal deposition parameters, choosing the appropriate deposition techniques and the post-deposition annealing processes required to remove process-induced (extrinsic) defects [1].

2.2.1 Accelerated Degradation Tests

Traditional reliability analysis involves the use of life data obtained from tests carried out under normal operating conditions. In many instances, such life data is very difficult, if not impossible to obtain especially when the products are designed to operate without failure for several years [33]. For instance, transistors are often expected to last for up to 10 to 20 years without failure consequently, very few failure data points would be available (if tested under normal operating conditions) to actu-

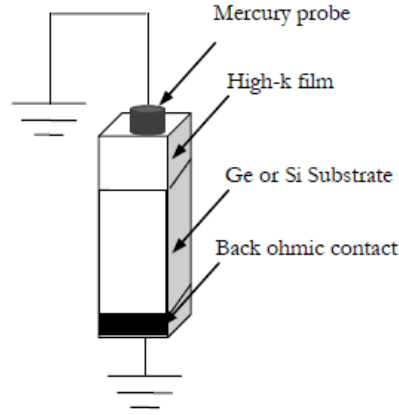


Figure 2.3. Schematic representation of a High-k MOS Capacitor (MOSCAP) structure for mercury probe measurements

ally reflect the life profile of its components, including the dielectric film. It is thus a significant challenge to obtain sufficient data to estimate the underlying probability structure of such highly reliable materials under normal operating conditions.

Given this difficulty, and the inescapable need to observe product failures in order to better understand their failure modes and underlying life characteristics, a variety of approaches have been devised to induce product failure more quickly than would occur under normal use conditions. The methods and procedures that are used to accelerate failure occurrences are known as accelerated life tests (ALT) [34]. In general, accelerated life tests (ALT's) involve over stressing the product by exposing it to severe conditions than those experienced in normal conditions. These severe stresses accelerate product failure occurrences for the purpose of quantifying the life characteristics of the product at normal use conditions [35]. Often, lower accelerated test conditions closest to the normal use conditions are used to minimize dielectric film reliability projection errors [43].

Dielectric failure is never sudden but rather its insulation ability gradually degrades until it reaches a critical value, usually a percentage of the initial performance level at time $t = 0$. As a result, an alternative approach to assessing dielectric reliability is monitoring the degradation of its characteristics of interest such as its performance which is indicated by the progressive increase. This approach is known as Accelerated Degradation Testing (ADT). Accelerated Degradation Tests ADTs measure product performance to determine the point at which performance degradation sets in, thus leading to product failure [34].

The concept of reliability analysis using data from accelerated tests is based on the following premises and assumptions [43]: (i) A component operating under elevated stress levels will exhibit the same mode of failure, as at normal use conditions [34]. In other words, the shape of the underlying failure distribution does not change, but the location of the distribution may change from one stress level to another. (ii) The variations inherent in the resulting test data are due to the random statistical fabrications and test process variations. This assumption can carefully be met by appropriately designing the test experiments, ensuring quality and repeatable fabrication and using well calibrated test equipment. (iii) The samples stressed are representative of the population. (iv) Each sample fails independently. (v) It is possible to represent the discrete failure events with continuous functions.

In this study, we will assume that in the ALT and ADT (both referred to as Accelerated Failure Tests (AFT)) models, the applied stresses (covariates) act multiplicatively on the failure times (linear AFT models) [44]. Also, we will assume that the applied stresses are within the acceptable region of true-acceleration [5].

Tobias and Trindale [45] state that true acceleration takes place when every failure time and every distribution percentile at the test condition is proportional to the projected results at the operating condition by a similar proportionality constant,

thereby resulting in a linear acceleration. When true acceleration occurs, the resultant reliability inference measures such as the time to failure $t_{(.)}$, the probability failure distribution $f_{(.)}$, the cumulative failure distribution $F_{(.)}$ and the failure rate $h_{(.)}$ at use conditions relate to the corresponding reliability measures at test conditions as follows:

$$\begin{aligned} t_{use} &= \mu t_{test} \\ F_{use}(t) &= F_{test}\left(\frac{t}{\mu}\right) \\ f_{use}(t) &= \left(\frac{1}{\mu}\right) f_{test}\left(\frac{t}{\mu}\right) \\ h_{use}(t) &= \left(\frac{1}{\mu}\right) h_{test}\left(\frac{t}{\mu}\right) \end{aligned}$$

where the subscripts *use* and *test* describe the operating and the test conditions and μ represent the proportionality constant also known as the acceleration factor.

CHAPTER 3

RESEARCH MOTIVATION AND RESEARCH QUESTIONS

3.1 Why Nanoreliability?

Nanotechnology has revolutionized virtually all industrial sectors through the introduction of new materials, design methods, fabrication, control, and measurements at the atomic level. To date, nanotechnology has resulted in the creation of high-performance materials and devices for mechanical, electrical, thermal, magnetic, and chemical applications. In addition, introduction of nano-bio-technology has had profound effects in pharmaceuticals and targeted drug delivery, which have enabled the eradication of illnesses through subcellular control and disease diagnostics [46]. The impact of nanotechnology has also been extended to the electronic industry where faster microprocessors are being developed and applied to advance supercomputing. In sustainable developments, nanotechnology has resulted in new methods of water reclamation, energy transformation and storage, and innovative agricultural systems [46]. Currently, the National Science Foundation envisions that nanotechnology has the potential to be a trillion dollar industry by 2015, as indicated in table 3.1. Figure 3.1 shows NSF's vision for the potential spending in nanotechnology by geographical region [47].

The National Academy of Engineering (NAE) and NSF recently proposed several 21st century science and engineering grand challenges among which is the advancement of tools of scientific discovery including nanotechnology. This study will con-

Table 3.1. Potential percentage value of nanotechnology per sector in 2015

Industrial Sector	Amount (\$Billion)	Percentage of Total Nanotechnology Spending
Materials	340	31%
Electronics	300	28%
Pharmaceuticals	180	17%
Chemical Manufacture	100	9%
Aerospace	70	6%
Sustainability	45	4%
Improved Healthcare	30	3%
Tools	20	2%
Total	1085	100%

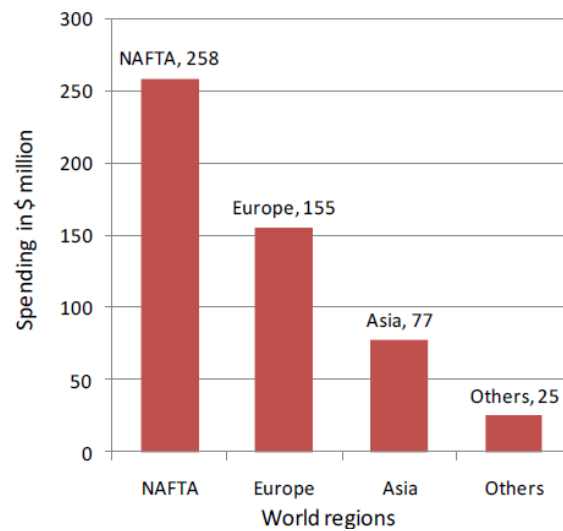


Figure 3.1. Development of nanotechnology markets worldwide by regions (NSF 2015 forecast)

tribute to the grand challenges by providing an understanding of the inherent failure modes and hence the reliability of metal/high-k gate stacks of Si-based substrates.

The successful fabrication and incorporation of metallic oxides in transistors was a major milestone and significant achievement in the electronics industry. With the increasing scaling of transistors, the gate dielectric film's thickness is predicted to progressively decrease to a record atomic size thickness. The failure of such thin gate dielectrics is based on quantum interaction of charge carriers, inherent film defects, and the applied stresses such as the electric field and the voltage across the films. The major driver in the reduction of transistor dimensions has been the need to increase gate capacitance and thereby increase current flow and device performance. However, the reduction in the transistor dimensions and more specifically gate dielectric thickness to about 2 nm has had severe reliability implications and challenges including leakage currents due to quantum-mechanical tunneling and conduction pathways formed by intrinsic defects, leading to increased power dissipation which causes temperature and time related dielectric degradation.

3.2 Nanoreliability Framework

In this study, we develop a framework for nanoreliability that incorporates methods of simulation and experiment to consistently describe dielectric failure with the aim of predicting their life characteristics, using appropriate statistical methods. Figure 3.2 shows a general reliability framework which can be extended and appropriately modified to study the reliability of not only other nanofilms but also to other nanomaterials and devices.

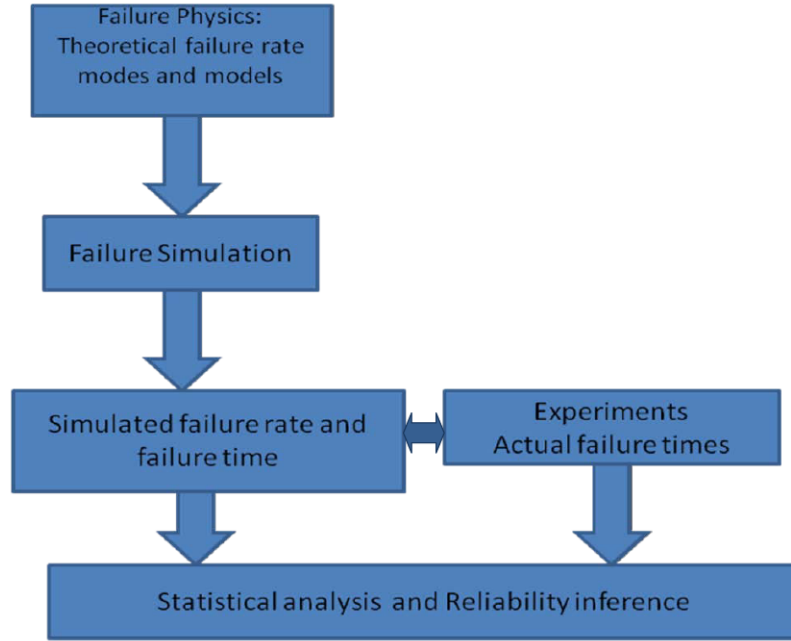


Figure 3.2. A general reliability framework

3.3 Statement of the Problem and Research Questions

In this study, we implement the reliability framework in addressing the following challenges, that will deepen the understanding of dielectric failure.

1. Is there a relationship between defect generation and dielectric thickness?

Thickness reduction of SiO_2 dielectric films as CMOS technology advances has reached its limits due to leakage currents as a result of electron tunneling and stress-induced defects. This limitation, necessitates the introduction of alternative high-k dielectric films as replacements for SiO_2 , with the aim of allowing for thicker films and at the same time improving gate performance. By using a thicker film, direct tunneling of electrons is expected to be eliminated and the critical defect density required to trigger current leakage is expected to increase. Since the critical defect density is proportional to dielectric thickness, we expect that the Mean Time to Failure (MTTF) will increase in increasing film

thickness. While this conclusion may be intuitive, a question arises regarding the actual relationship between the film thickness and the time to dielectric breakdown. We propose to address this question by developing a dielectric simulation model whose results will give clarity to the actual relationship between defect generation, time to dielectric breakdown and dielectric thickness.

2. What is the profile of the dielectric failure density function?

Statistical analysis of failure data is very central to the study of material or device reliability and it concerns the analysis of the behavior of the material or device with time, especially during its useful life. Reliability engineers are typically concerned with the failure rate of a device. For instance, it is generally agreed that most electronic devices and systems exhibit failure rate behaviors that are divided into three stages, namely: (i) an initially decreasing *early life* failure rate also known as the burn-in period, (ii) a constant or useful life failure rate, and (iii) an increasing wear-out portion, also called the aging period [48]. With the development of nanomaterials and devices, questions about the form of their failure rate have arisen. In this study, we will seek to understand the structures of the hazard function, the failure density function and the corresponding reliability function of dielectric films. The structure of these functions provide important insight into the nature of the life characteristics of the device.

3. How do we project reliability at test conditions to normal use conditions?

The question regarding the appropriate extrapolation model arises when the failure data used in reliability analysis is acquired from experiments carried out under accelerated conditions [49]. As mentioned in chapter 2, accelerated degradation tests are useful for materials and devices that are designed to operate without failure for several years. In this case, the materials and devices

are tested by exposing them to severe stress conditions than they would experience under normal operating conditions. Since the aim of failure acceleration is to obtain enough data to quantify the life characteristics of the materials at normal use conditions, the question then is what extrapolation model would be appropriate to relate reliability inference at test conditions to those under normal operating conditions? In this study we will derive the extrapolation model from the physical failure model, which relates failure time to stress levels.

CHAPTER 4

SIMULATION OF DIELECTRIC BREAKDOWN

4.1 Modes of Dielectric Breakdown

In this chapter, we first present the two main dielectric failure mechanisms: electron quantum tunneling and stress-induced defect generation. Later we present the 3D failure model to simulate dielectric failure due to stress-induced defects, which is the focus of this work.

4.2 Current Leakage Due to Quantum Tunneling

In thin dielectric films, electrons can tunnel through the dielectric by the quantum tunneling mechanism, which involves the movement of electrons through a trapezoidal barrier caused by an energy band shift as represented in figure 4.1. Using the Wentzel-Kramers-Brillouini (WKB) approach, the tunneling probability P_ϵ exponentially increases as the dielectric thickness decreases, as seen from the following expression [1]:

$$P_\epsilon = \exp \left[\frac{-2i}{h} \int_{x_1}^{x_2} \sqrt{2m(\epsilon - U(x))} dx \right] \quad (4.1)$$

where $(x_2 - x_1)$ is the tunneling distance. The amount of tunneling current is determined by calculating the tunneling current density J_{tunnel} flowing through the high-k dielectric region using the following expression:

$$J_{tunnel} = \frac{(m_{ox}m_{hk})^{\frac{1}{2}}k_B T}{2\pi^2 h^3} \int_{E_F}^{\infty} P(E, V) \ln \left[\frac{1 + \exp[(E_F - E)/k_B T]}{1 + \exp[(E_F - E - qV_G)/k_B T]} \right] dE \quad (4.2)$$

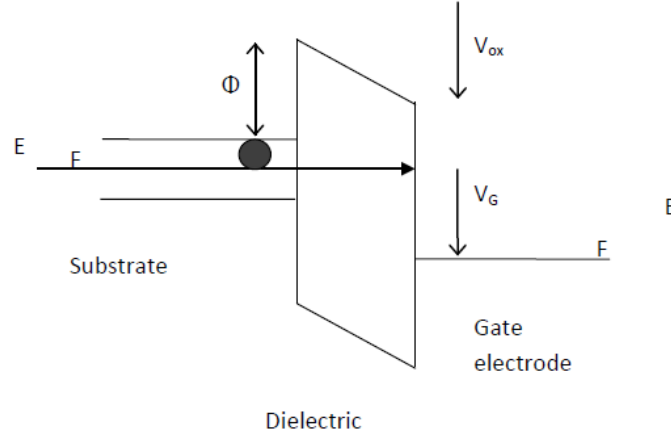


Figure 4.1. A schematic energy band diagram of the MOS structure where Φ is the energy barrier height at the dielectric/substrate interface, V_{ox} is the potential drop across dielectric film and V_G is the applied gate voltage [1]

where m_{ox} and m_{hk} are the electron tunnel effective mass in the interfacial oxide and bulk high-k layers respectively, and $P(E, V)$ is the electron tunneling probability.

To avoid electron tunneling, the barrier band or band offset should be over 1 eV for both conduction and valence bands [1]. Even though SiO_2 has a wider band gap than almost all high-k dielectrics as was shown in table 1.2, the high-k band gaps are aligned with the Si substrate in such a way that their conduction band offset are smaller than the valence band offset. It turns out that this differential band offset limits current leakage in high-k films [1]. In this study we will only consider dielectric failure as a result of stress-induced defect generation.

4.3 Stress-Induced Defect Generation: Physical Models

Even though dielectric failure is quite complex, it is widely accepted that dielectric degradation occurs mainly as a result of stress-induced defect generation within the dielectric bulk and at the interface regions [32]. These defects include neutral electron traps, hole traps, and interface states [27]. The random nature of the breakdown

process and its dependence on atomic structure, temperature, stress and dielectric thickness has led to the concepts of multi-modal failure distributions. These distributions are linked to the intrinsic and extrinsic failure modes described earlier.

Extrinsic failures are caused by defects formed during the fabrication and packaging processes, whereas intrinsic failures are caused by stress-induced defects also referred to as charge traps within a would-be perfect dielectric film. As technologies mature, many electronic fabrication and packaging companies are registering high quality yields, which means that process-induced extrinsic failures are progressively being reduced [50]. This leaves out the intrinsic stress-induced defects as the main cause of defect breakdown [27]. Currently, there are two physical models that explain stress-induced generation of defects, namely the electron-energy dissipation model and the thermo-chemical degradation model.

4.3.1 Electron-Energy Dissipation Model (1/E model)

The energy dissipation model, also known as the Anode-Hole injection (AHI) model suggests that as charge carriers channel from the source to the drain during the inversion (ON) state, they release some energy upon collision with the anode material through the process of impact ionization [51]. The energy that is released is enough to break the hydrogen bonds at the dielectric/substrate interface, and therefore causes hydrogen ion release at the anode region. The released hydrogen ions penetrate the dielectric causing lattice mismatch, and therefore inducing defect sites where charges are trapped. Figure 4.2 illustrates the release of hydrogen ions which penetrates the dielectric film. The AHI model assumes that the damage in the dielectric is proportional to the hole fluence (density) and that the lifetime of the gate dielectric is determined by the time before the hole fluence Q_p reaches some critical value [27]. Even though it is difficult to measure hole fluence, the density

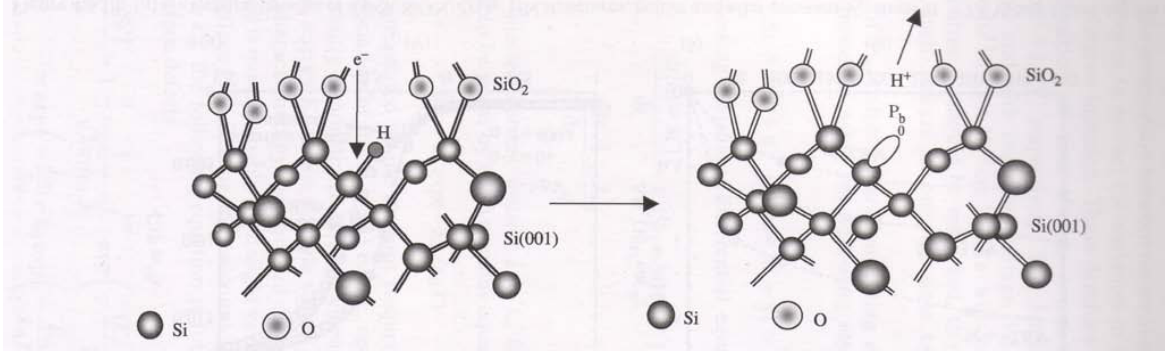


Figure 4.2. Generation of a Pb trapped charge defect after impact ionization [1]

can be approximated by measuring the electron energy lost during impact, especially because the concentration of holes can be equated to the number of broken bridging oxygen bonds [1]. The hole fluence Q_p is expressed as follows:

$$Q_p \propto JRt \quad (4.3)$$

where $J \propto e^{-B/E_{diel}}$ and $R \propto e^{-H/E_{diel}}$. Therefore equation 4.3 becomes:

$$Q_p \propto e^{-\frac{(B+H)}{E_{diel}}} t \quad (4.4)$$

In equations 4.3 and 4.3, J and R denote the hole-generation coefficient and the current density respectively, and t is the stress test duration. B and H are process constants in MV/cm and E_{diel} is the electric field across the dielectric. According to the AHI model, the rate of dielectric breakdown τ_{bd} is proportional to the hole fluence Q_p , and is modeled as follows [52]:

$$\tau_{bd} \propto \exp - \left[\frac{B + H}{E_{diel}} \right] \quad (4.5)$$

Equation 4.5 was modified to incorporate the effect of defect growth also referred to as the effective dielectric thinning, as an important parameter in the breakdown model. The modified anode-hole injection model is given as [52]:

$$\tau_{bd} \propto \exp - \left[\frac{(B + H)(X_{diel} - \Delta X_{diel})}{E_{diel}} \right] \quad (4.6)$$

where $(X_{diel} - \Delta X_{diel})$ is the resultant dielectric thickness due to the effective thinning caused by increased defect dimension (ΔX_{diel}) .

4.3.2 The Thermo-Chemical Model (E-model)

The thermo-chemical model suggests that defects are generated as a result of a thermo-chemical reaction that is driven by the applied electric field and the stress time. The electric field within the dielectric causes thermal weakening or stretching of the bridging oxygen bonds thereby lowering the activation energy required to break the bonds. The broken bonds generate voids that eventually trap charges thus creating defects in the dielectric bulk and at the interface as shown in figure 4.3.

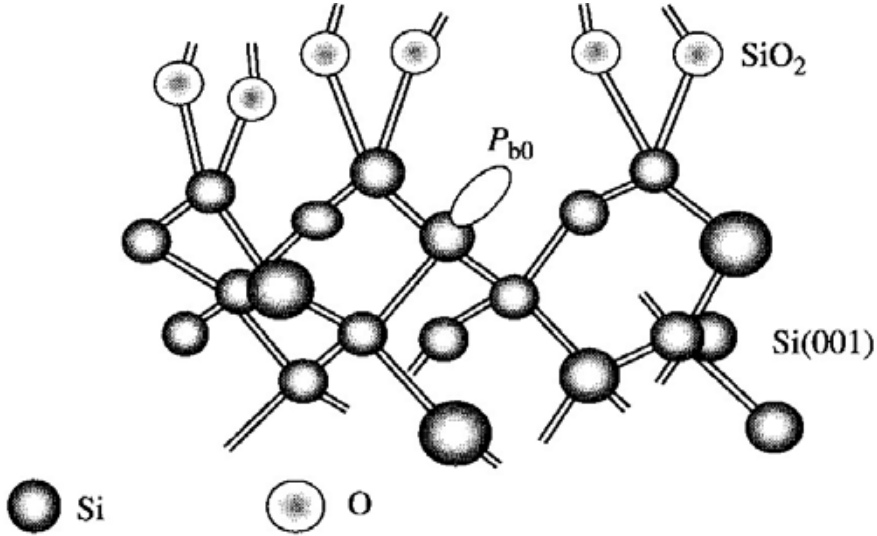


Figure 4.3. Generation of a Pb charged defect by the thermal-chemical process [1]

This model is based upon the Arrhenius model which is typically used to express the rate τ , of most thermo-chemical reactions as a function of the reaction parameters. When applied to dielectric failure, the Arrhenius breakdown formula indicates that the rate of defect formation τ_{bd} is exponentially related to the applied electric field E_{diel} as follows [27]:

$$\tau_{bd} \propto \exp - \left[\frac{\Delta H}{k_b T} - \gamma E_{diel} \right] \quad (4.7)$$

where: ΔH is the activation energy in eV, k_b is the Boltzmann constant in eV/K, γ is the material dependent field acceleration parameter in cm/MV, T is the absolute temperature in Kelvin and E_{diel} is the dielectric electric field in MV/cm [52].

There is debate among researchers concerning the acceptability of the two physical failure models [27], [53], [54]. For instance, the anode-hole model has been criticized for having some limitations. According to McPherson et al. [52], the AHI model defies the concept of thermal diffusion which states that all material degrades even in the absence of an electric field. Thermal diffusion occurs when increased operation temperatures cause impurities to diffuse into the dielectric/substrate interface, causing a change in the structure of the depletion channel, hence reducing transistor performance. Following suggestions by McPherson [52], the breakdown simulation in this study will specifically involve the thermo-chemical model.

Regardless of the use of the either model, researchers agree that dielectric breakdown is triggered when sufficient traps build up inside the dielectric, causing local conduction pathways that lead to current leakage [15]. All the models are capable of accommodating the sub-linear trap generation dependence on electron density, if the traps are proportional to the number of broken bonds [3]. The number of broken bonds can be determined from the defect density. McPherson et al. suggest that thick film breakdown is explained by the 1/E-model, particularly at higher fields.

On the other hand, thin film breakdown is suggested to follow the E-model especially under lower electric field level [27]. Much effort has been devoted to determine the temperature and field acceleration factors especially when the failure tests are carried out in elevated temperatures and electric field in order to obtain a substantial amount of failure data [55].

4.3.3 Defects Identification and Quantification

The dielectric could be considered perfect in the absence of broken-bonds, voids, and impurities. Gibson and Dong in Houssa et al. [1] suggest the presence of voids of 1 nm in diameter in hf-based high-k films. The number and distribution of such voids depend on the stress duration, the applied electric field and test ambient conditions. Stemans and Afanas in Houssa et al. [1] developed a complementary approach for atomic defect identification and quantification using the electron spin resonance (ESR) method.

ESR spectroscopy is a technique for studying chemical species that have one or more unpaired electrons, by applying a magnetic field to the samples containing the charged species. Currently, most commercially available high-k-on-Si stacks either use HfO_2 or ZrO_2 as the high-k material in their pure or silicate forms [1]. Previous research has shown that the defects present in Si/SiO₂ stacks are in the paramagnetic Pb-type charge centers and they also appear to be the most prevalent defect type in Si/high-k stacks [1].

Current advanced ESR spectrometers can detect up to $1 \times 10^{13} \text{ cm}^{-2}$ defects at low temperatures within acceptable signal averaging times [1]. As an example, Houssa et al. studied $\text{Si}/\text{SiO}_x/\text{ZrO}_2$ and $\text{Si}/\text{Al}_2\text{O}_3/\text{ZrO}_2$ stacks prepared on low-doped one-side (100) Si wafers. Uniform stoichiometric layers of Al_2O_3 and SiO_2 (0.5 to 3nm) followed by ZrO_2 and HfO_2 (5 to 20 nm) were deposited in an atomic layer

deposition (ALD) reactor. Samples of 2 by 9 mm² were tested using ESR and their results showed that spin densities of $5.9 \pm 0.3 \times 10^{12} \text{cm}^{-2}$ and $1.5 \pm 0.2 \times 10^{12} \text{cm}^{-2}$ were recorded for the $\text{Si}/\text{SiO}_x/\text{ZrO}_2$ and $\text{Si}/\text{Al}_2\text{O}_3/\text{ZrO}_2$ stacks respectively. We faced challenges as we sought to acquire spacial defect measurements, so we will not report any actual spacial defect distribution and density. Instead we will proceed to describe the 3D simulation model in the next section.

4.4 Defect-Based 3D Breakdown Simulation Model

From the thermo-chemical model presented in section 4.3.2, the rate of defect generation increases exponentially with increasing electric fields. Since the rate of defect generation is inversely proportional to the time to defect generation, the time to dielectric breakdown can be inferred to exponentially decrease with increasing electric field. Also, Cheung [32] indicates that the breakdown probability depends on the gate dimensions and so does the defect density at breakdown. So far, several attempts have been made to simulate dielectric failure. Degraeve et al. [26], [56] and Stathis et al. [57] presented the first 2D percolation models, which they used to successfully explain the scaling of breakdown distribution with gate area. These percolation models involve a random generation of spherical defects within the simulation model, with the aim of getting the defect density at the point of dielectric breakdown in order to statistically determined the parameters of the underlying failure distribution. Subsequently, other researchers extended the 2D model to a 3D model, where the defects were generated following a Poisson distribution [3], [14], [25], [53] and [58].

In these 3D models, failure occurs when a critical number of defect density in the dielectric volume is reached. These existing dielectric failure models share the assumption that the underlying failure distribution is known to be Weibull, and this assumption stems from the premise that dielectric failure follows the weakest link

argument. The weakest link argument was introduced by Waloddi Weibull [59], [60], and suggests that: (i) failure occurs following the presence of the first instance of a critical amount of flaws, in our case, conduction paths, and (ii) the defects are uniformly distributed throughout the dielectric layer [61].

In this study, we adopt the 3D simulation model presented by Sune et al. [3] and use it to simulate dielectric failure and thereby generate simulated dielectric failure time data. Two key differences between our work and that presented by Sune et al. are: (i) While Sune et al. equate the rate of breakdown τ_{bd} presented in equation 4.5 to the rate at which a conduction pathway is formed, we equate τ_{bd} to the rate at which one defect is formed. This contribution will be clarified in the simulation model. (ii) While the Weibull distribution is most fitting for failure phenomena that are described by the weakest link argument, making assumptions regarding the distribution underlying the failure data might lead to misleading reliability inference. This contribution will be discussed in detail in chapter 5 and 6 of this study.

4.4.1 3D Breakdown Simulation Model Details

Assume that the dielectric film which consists of a cross sectional area A_{diel} is segmented into several a_0 cubic lattice cells as shown in figure 4.4 [53], [62]. Breakdown is triggered when the number of defects N_t , is high enough to form a critical number of conduction paths that traverse the dielectric thickness. The conduction path is assumed to be formed when the number of defects in a column of cells as seen in figure 4.5 reaches or exceeds the threshold value n_{bd} expressed by [63]:

$$n_{bd} = \frac{t_{diel}}{a_0} \quad (4.8)$$

where t_{diel} is the dielectric thickness and a_0 is the cubic cell dimension. Though the conduction path may not always be vertical, n_{bd} is the least number of defects required to form a conduction path.

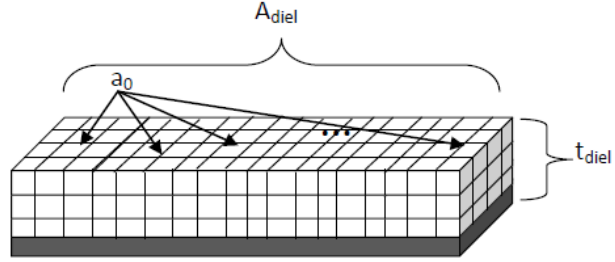


Figure 4.4. Division of the dielectric in a lattice of 3D cells with lattice constant of a_0^3

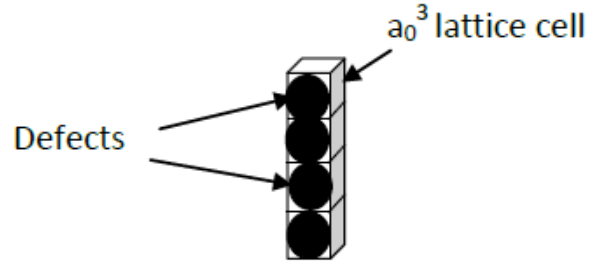


Figure 4.5. Point defect generation forming a conduction path

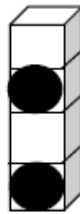


Figure 4.6. Point defects are insufficient to form a conduction path

According to the weakest link theory, the dielectric film breaks down if the number of defects in at least one the columns reaches or exceeds n_{bd} , and is therefore enough to form a conduction path. There has been questions as to whether the formation of one filament is enough to trigger a hard breakdown. To this effect, we adopt the suggestion by McPherson et al., that while the three dielectric failure modes, namely soft breakdown, stress-induced leakage currents (SILC) and hard breakdown are related and occur at different times. They are also cumulative, which means that soft and SILC breakdown, progressively lead to hard breakdown [52]. McPherson also suggests that while the dielectric might still be functional under soft and SILC breakdown modes, severe degradation is still taking place within the dielectric with time [52]. To this effect, we set the simulation to indicate dielectric failure when five or more conduction paths are formed.

In this study, we consider defect generation to be field-driven and to follow a homogeneous Poisson process [14]. We also present a quantitative relationship between defect generation and dielectric time to breakdown. In this failure simulation model, two variables which capture the possibility of breakdown change with area scaling are n_{bd} and a_0 , in which case n_{bd} varies with dielectric thickness.

Let the total number of defects at breakdown be N_t . The defect density, N_{diel} in defects/ nm^3 is given by:

$$N_{diel} = \frac{N_t}{A_{diel}t_{diel}} \quad (4.9)$$

and the fraction of defective cells, λ in the simulation is given by:

$$\lambda = \frac{N_t a_0^3}{A_{diel}t_{diel}} \quad (4.10)$$

where a_0 is the cubic size, A_{diel} is the cross sectional area of the entire dielectric and t_{diel} is the dielectric thickness.

Let the probability that a cell will have a defect be estimated by the fraction of defective cells is λ as follows:

$$F_{cell}(\lambda) = \lambda \quad (4.11)$$

The cumulative failure probability of the column is given by:

$$F_{col}(\lambda) = (F_{cell}\lambda)^{n_{bd}} = \lambda^{n_{bd}} \quad (4.12)$$

Using the 3D cell model and following the weakest link theory, the thin film will fail whenever a conduction path forms in any of the columns. Therefore, the probability that the dielectric will be functional (insulating) is given by the probability that all the columns in the simulation are insulating, meaning that none has a conduction path. Therefore the reliability of the dielectric is expressed as:

$$R_{diel}(\lambda) = (1 - F_{col}(\lambda))^{N_{col}} = (1 - \lambda^{n_{bd}}) \quad (4.13)$$

Here, $N_{col} = \frac{A_{diel}}{a_0^2}$ is the number of columns in the 3D model.

So far, equations 4.11 to 4.13 relate the dielectric reliability to the probability of generating one defect λ . However, the aim of this simulation model is to derive simulated failure times t_{bd} from the rate of defect generation. Using the appropriate defect density measurement techniques such as Electron Spin Resonance (ESR) spectroscopy, the time t_{bd} , that it takes to generate the critical number of defects N_t , for breakdown to occur can be monitored in real time, and used to calculate the rate at which defects are generated which will be denoted by ν in defects per seconds. With this rate, and assuming that defect generation follows a homogeneous Poisson process, the probability that a column of the 3D model will not fail R_{col} is given by the probability of generating less than n_{bd} defects in the column as follows:

$$R_{col} = \sum_{k=1}^{n_{bd}-1} \frac{\nu^k e^{-\nu}}{k!} \quad (4.14)$$

Let the number of columns in a dielectric be given by N_{col} . Following the weakest link approach, the failure probability of the dielectric is equal to the probability that one or more columns will fail [63], [64], [65], [66]. Alternatively, the reliability of the dielectric R_{diel} can be determined from equation 4.14 as the probability that none of the columns will fail as follows:

$$R_{diel} = [R_{col}]^{N_{col}} \quad (4.15)$$

$$R_{diel} = \left[\sum_{k=1}^{n_{bd}-1} \frac{\nu^k e^{-\nu}}{k!} \right]^{N_{col}} \quad (4.16)$$

4.4.1.1 3D Simulation Model Assumptions

The following assumptions were used in setting up the failure simulation model:

1. A defect is spherical with diameter $d=1$ nm [3], [66].
2. One defect fills a cubic cell as seen in figure 4.7.

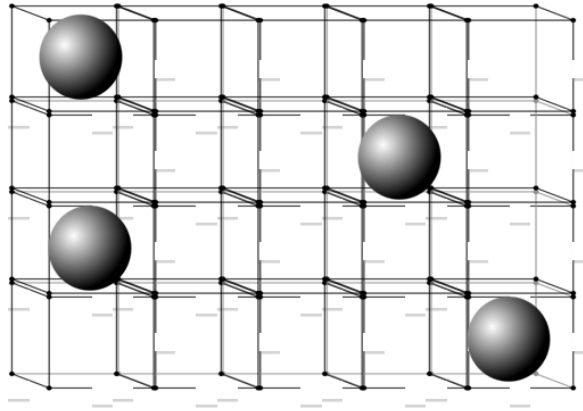


Figure 4.7. 3D representation showing some defective cells

3. Each cell has 6, 18 or 26-degrees of communication with neighboring cells, that is, 6, 18 or 26-nearest neighbor cells.
4. Defects are introduced into the simulation model following a homogeneous Poisson process.
5. Cells have equal probability of being filled by defects. This implies that atomic bonds within the dielectric bulk and at the interface regions have equal probability of being broken [1].
6. Two defects cannot occupy the same cell, which means that once a cell is occupied by a defect it cannot be occupied by another defect.
7. The dielectric 3D representation is declared defective when at least five conduction paths are found from the top (gate electrode) to the bottom (semiconductor substrate). A comparison is later made between the results obtained when the required number of conduction paths is greater than one.

4.4.1.2 3D Simulation Model Pseudocode

The following is the pseudocode of the steps that were followed in setting up the failure simulation model:

1. Represent a defect-free dielectric by a 3D array of size L by W by H , where (L , W and H) represent the number of cells in the x, y and z directions, equivalent to the dielectric volume $A \text{ nm}^2$ by t_{diel} nm (see figure 4.8). For this simulation, $L=45$ nm represents the gate channel, $W=90$ nm represents the gate width, and $H=t_{diel}$ represents the dielectric thickness.
2. Randomly introduce defects according to a Poisson distribution with a mean value of λ .

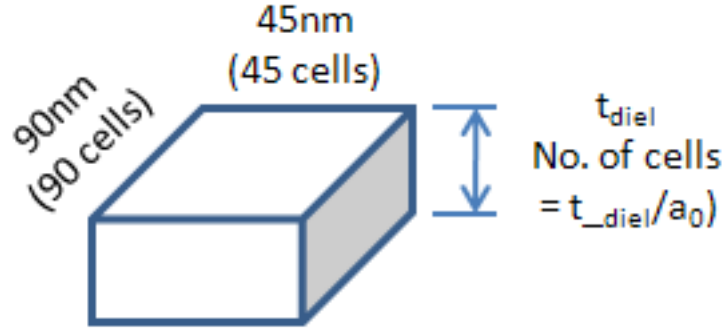


Figure 4.8. 3D simulation size

3. Introduce the defects into the array uniformly. Figure 4.9 indicates a 2D representation of defect filled cells represented by 1s.



Figure 4.9. 2D representation showing defect filled cells as 1s, and defect-free cells as 0s

4. Find clusters of communicating cells as shown in 2D in figure 4.10.
5. Search for a conduction path by searching for individual clusters that bridges the 3D array.
6. Declare the 3D array (simulated dielectric) failed (conducting) when the critical number of conduction paths are found [62], [63], [64]. We include simulations for at least five conduction paths to address concerns that a single conduction path may not necessarily cause failure.

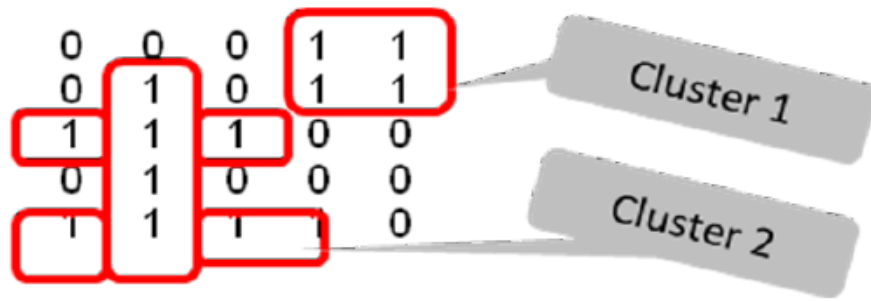


Figure 4.10. 2D representation showing clusters of communicating cells

7. If the simulation array has not failed, declare the array functional (insulating)
repeat the simulation steps (starting from step 2).

Figure 4.11 is a 2D representation of an insulating array, and figure 4.12 is a 2D representation of a failed array.



Figure 4.11. 2D representation showing an insulating array

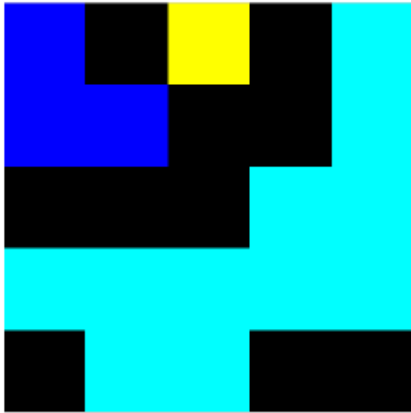


Figure 4.12. 2D representation showing a failed (conducting) array

4.5 Simulation Results

In the simulation model, we assume a constant cross sectional area of 45 by 90 nm^2 as shown in figure 4.8 with the thickness varying from 1 to 5 nm. At each thickness, 1000 replicates are carried out, each replicate representing a dielectric sample. The defect diameter which is equal to the cubic cell dimension, a_0 is fixed at 1nm [3]. Figure 4.13 graphically shows the relationship between the critical defect density at breakdown and the dielectric thickness when 6-degrees of communication (face-oriented), 18-degrees, and 26-degrees of communication (face and edge-oriented) are considered. The figure shows an increasing, non-linear relationship between the critical defect density and dielectric thickness. This is an indication that it takes more defects to form as many conductive paths in thick films than in thin films, for a fixed dielectric cross sectional area.

To validate our simulation results, the breakdown defect densities at given dielectric thickness were compared to the analytical model proposed by Sune et al. [67], which relates defect density at breakdown to the simulated dielectric thickness as follows:

$$N_{bd} = \frac{t_{diel}}{a_0^3} \exp \left[-\frac{1}{\beta} \ln \left(\frac{A_{diel}}{a_0^2} \right) \right] \quad (4.17)$$

where β is the Weibull shape parameter is approximated by:

$$\beta = \frac{\alpha(t_{int} + t_{diel})}{a_0} \quad (4.18)$$

The interface oxide thickness, t_{int} in equation 4.18 is assumed to be 0.37 nm and α is a proportionality constant given as 0.38 in [62], [64]. The defect density from the

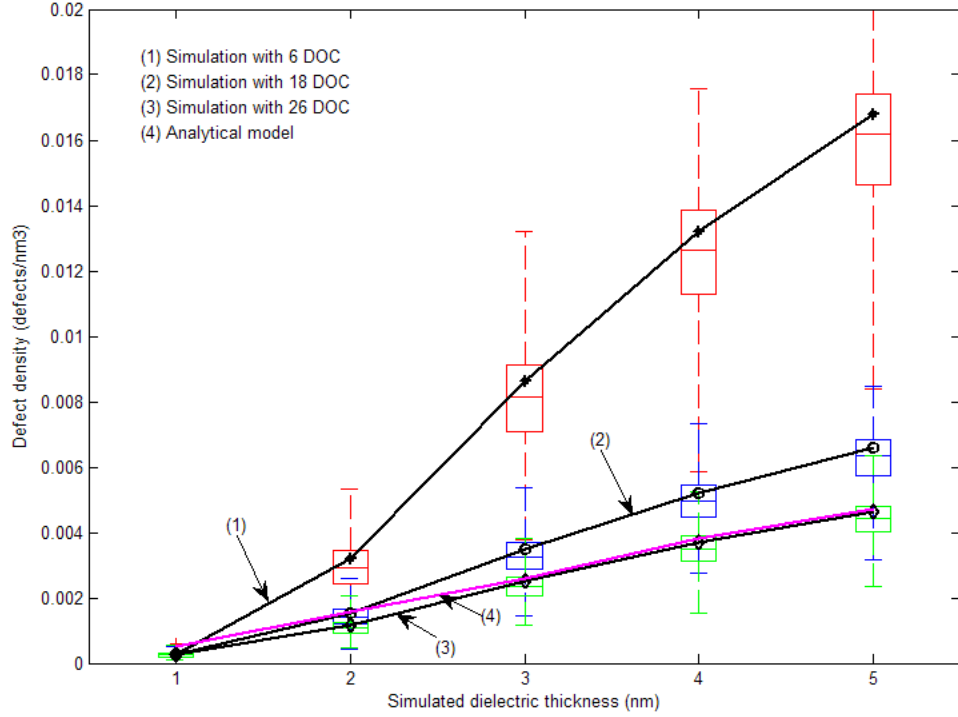


Figure 4.13. Critical defects densities for 1, 2, 3, 4 & 5 nm dielectric thickness

analytical model is shown in figure 4.13 and the plots show that the defect density based on 26 degrees-of-communication (DOC) is closest to the analytical model.

We further analyzed the effect of the number of critical paths and mean number defect (the average number of defects introduced in the simulation model) generated at each iteration on the critical defect density for each simulated thickness. The plots in Figure 4.14 show that the critical defect density when the critical number of paths is at least five is closest to the analytical model. The figure also shows the critical defect density when the mean number of defects introduced at each iteration was decreased from 10 (line plot 3) to 1 (line plot 4) are quite similar. From the physics of material point of view, this probably means that the effect of increasing the rate of defect generation is less in thin dielectrics ~ 1 to 5 nm. Unfortunately, this claim can

only be validated experimentally by spatially monitoring actual defect generation in a variety of dielectric thickness.

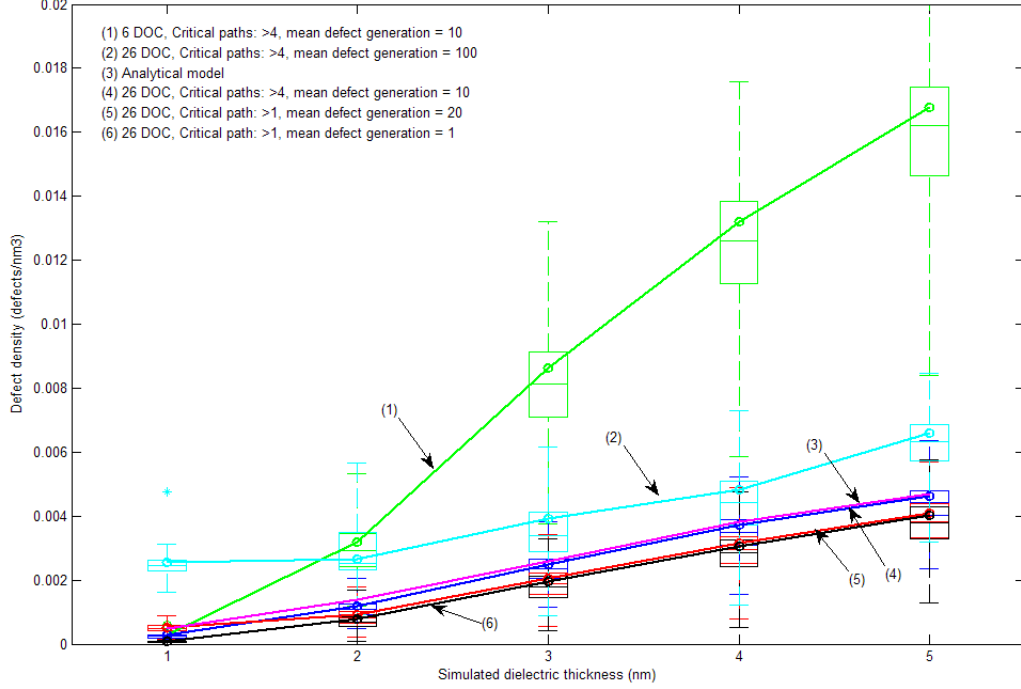


Figure 4.14. Effect of the critical number of paths on critical defect density

4.5.1 3D Simulation Model Extension to the Breakdown Time of the Dielectric

The aim of the 3D simulation is to relate the critical defect density to the breakdown time. To do so, we use the empirical defect generation time model proposed by Tous et al., which describes the relationship between time to dielectric breakdown, t_{bd} and the defect density, N_{diel} as follows [64], [65]:

$$N_{diel} = \xi t_{bd}^{\frac{1}{\alpha}} \quad (4.19)$$

where α is a constant, which in our simulation is maintained at 0.38 for all dielectric thickness. ξ is the defect generation efficiency defined as the number of defects generated for every charge carrier that is injected into the dielectric. ξ is therefore a function of the breakdown charge, Q_{bd} and it is expressed as [64], [65], [67]:

$$\xi = \frac{q(t_{diel})}{a_0^3 Q_{bd}} \exp \left[-\frac{1}{\beta} \ln(N_{col}) \right] \quad (4.20)$$

In equation 4.20, q is the absolute value of the charge carried by an electron and it is approximately equal to 1.602×10^{19} Coulombs. In their study, Sune et al. report constant values of ξ at different gate voltages as follows: (i) $\log(\xi) = -20$ for $V_G = 2.3V$, (ii) $\log(\xi) = -13$ for $V_G = 3.8V$, and (iii) $\log(\xi) = -8$ for $V_G = 5V$ [65], [67]. We used these values of ξ to generate simulated failure times corresponding to 2.3 V, 3.8V and 5V. Figures 4.15, 4.16 and 4.17 are box plots that graphically show the variability of the failure times for each simulated dielectric thickness (1 to 5 nm), when the gate voltage is assumed to be 2.3V, 3.8V and 5V respectively.

The figures show that the variability of the simulated failure data is consistent at each dielectric thickness and for each voltage level. Also, the simulated breakdown times increase with increasing gate voltage, which is in agreement with the assumption of the Arrhenius acceleration model.

4.6 Conclusion

In this chapter, we presented a 3D dielectric failure simulation model, in which the simulated dielectric film was represented by a 3D array. The length and width of the array correspond to the simulated gate dimensions, and the thickness correspond to the simulated dielectric thickness in nm. In the simulation, defects are randomly introduced into the simulated dielectric film and a search algorithm is used to monitor

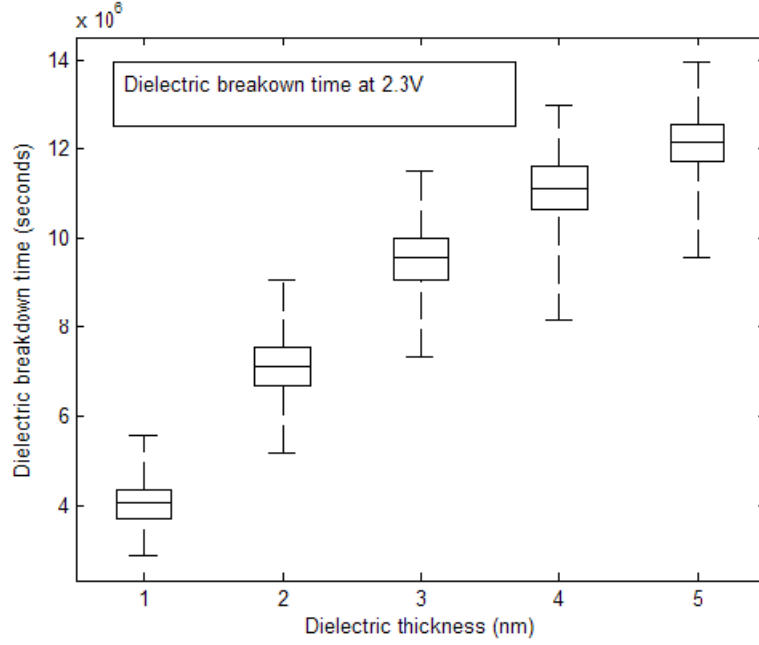


Figure 4.15. Simulation failure times at 2.3 V

formation of conductive paths. A conductive path is defined as a cluster of communicating cells which traverse the thickness of the dielectric. We varied the search algorithm by considering 6, 18 and 26-nearest neighbor cells, and also by defining breakdown as the point when at least five conductive paths are formed. For validation, we compared our results with an analytical dielectric failure model proposed by Sune et al. [67] that relates the critical defect density at breakdown with the simulated dielectric thickness. We also used a power law function that relates critical defect density at breakdown to the stress time to generate dielectric failure times.

Based on the simulation results we can conclude that:

1. The critical defect density increases with dielectric thickness.
2. It would appear that the path formation that uses 26-nearest neighbor cells most closely approximates the actual dielectric failure process.

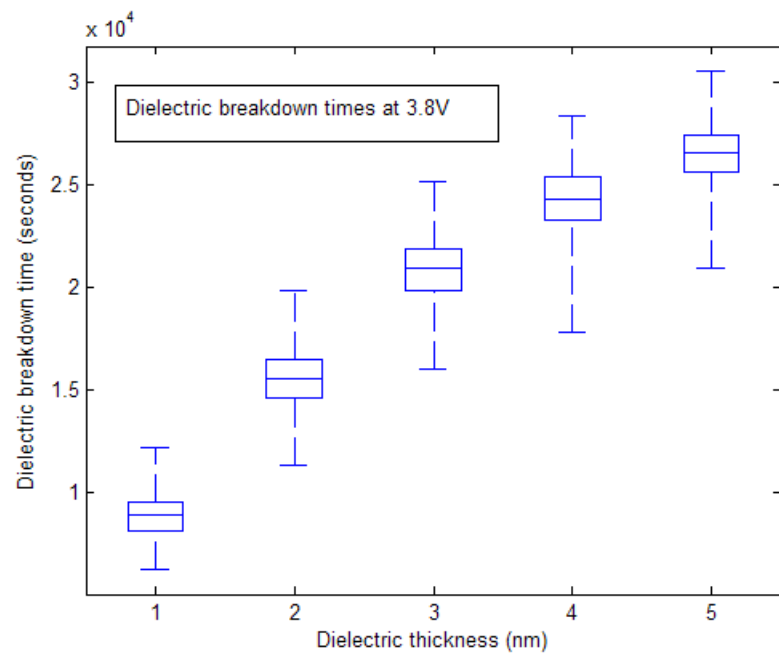


Figure 4.16. Simulation failure times at 3.8 V

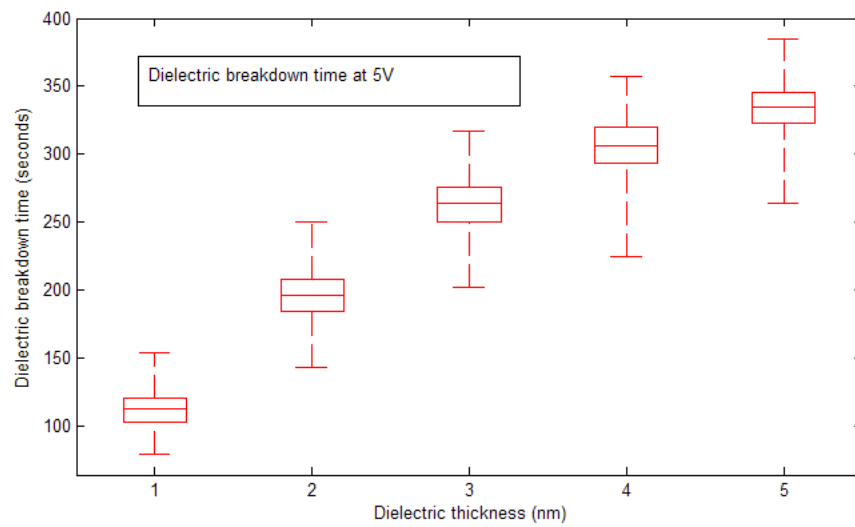


Figure 4.17. Simulation failure times at 5V

3. The mean number of defects introduced at each iteration largely affects the critical defect density. For instance remarkably high defect densities ranging from 0.015 to 0.05 defects/nm³ were realized when the mean number of defects introduced into the model was fixed at 100 for the 1 nm and 5 nm simulated dielectric thickness respectively. This range is high, compared to the range of 0.00007 to 0.004 when the mean number of defects was fixed at 1.
4. The variability of failure times is consistent at each simulated dielectric thickness. This is an indication that the simulation is quite robust in predicting the relationship of failure times to dielectric thickness.
5. The failure times increase with increasing thickness and gate voltage.

CHAPTER 5

NONPARAMETRIC ESTIMATION OF FAILURE TIME DISTRIBUTION: USING KERNEL DENSITY METHODOLOGY

In this chapter, we discuss the use of kernel density approach to construct the probability density structure of actual dielectric failure data, as well as the choice of important parameters, which include the kernel and the optimal bandwidth.

5.1 Background

Consider a continuous random variable X whose distribution follows a probability density function $f(x)$. The probability density of a given random variable X is the probability that its value belongs to a measurable region B of real numbers. Thus,

$$P(X \in B) = \int_B f(x) dx \quad (5.1)$$

if $B = [a, b]$ then we get:

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (5.2)$$

To determine the density function $f(x)$, n independent outcomes X_1, X_2, \dots, X_n of X are used to construct the estimate $\hat{f}(x)$. Density estimation methods such as histograms, kernel, and wavelet techniques are important in applied and theoretical statistics because they provide data analysts with a graphical view of the shape of the distribution [68].

Let x_1, x_2, \dots, x_n be independent and identically distributed (IID) random sample of the random variable X taken from the population Ω , where n is the sample size. The question is to determine the probability distribution functions of X based upon the observed values. The two main estimation approaches include: parametric and nonparametric methods. In parametric methods, it is assumed that the forms of the distributions are known, and corresponding unknown distribution parameters are estimated using the observations. Statistical inference techniques such as goodness-of-fit tests are used to determine if the underlying distribution belongs to a known classical distribution.

In reliability analysis, some of the common distributions include the exponential, Weibull, gamma and log-normal distributions, in which case we need only to estimate the associated parameters [69]. The drawback of parametric analysis is that the limited number of known distributions and parameters do not necessarily fit all of the function densities encountered in practice. In addition, most common distribution densities are unimodal, whereas most practical distributions are multi-modal. The use of a wrong probability distribution function leads to misleading inference conclusions.

In those situations where there are no known underlying structures for the data, nonparametric approaches are most useful. In nonparametric density estimation schemes, rather than restrict the possible forms of the underlying probability structure, we only need to impose some mild assumptions such as smoothness of the underlying probability structure. Therefore, nonparametric estimation approaches are robust to varied data structures and are able to uncover structural data features that typical parametric methods cannot reveal. Some nonparametric distribution estimation techniques include kernel estimates, orthonormal series approximations, maximum penalized likelihood estimates, smoothing splines, and wavelet estimates [68].

Kernel estimation is an unsupervised learning approach [70], and is based upon the Naive Density Estimate (NDE) which similar to the histogram, estimates the probability density at a point x by the proportion of samples that fall within a neighborhood of x defined by $x - \frac{h}{2}, x + \frac{h}{2}$ limits. Given n IID observations $x_1, x_2, \dots, x_n \in X$, the NDE is expressed as:

$$\hat{f}(x) = \lim_{h \rightarrow 0} (2h)^{-1} P(x - \frac{h}{2} < X < x + \frac{h}{2}) \quad (5.3)$$

The NDE $\hat{f}(x)$ is determined by choosing a small h such that:

$$\hat{f}(x) = \frac{x_i \in \mathcal{N}}{nh} \forall i \in [1, 2, \dots, n] \quad (5.4)$$

where \mathcal{N} is a small neighborhood around x defined by $(x - \frac{h}{2}, x + \frac{h}{2})$.

Let $\omega(x)$ be a weight factor at point x . The NDE at x can be re-written as:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \omega\left(\frac{x - X_i}{h}\right) \quad (5.5)$$

Similar to histograms, NDEs are not satisfactory for density estimation because they suffer from jumps at every $X_i \pm \frac{h}{2} \forall i = 1, 2, \dots, n$ points and therefore they result in unsmoothed density structures [71]. Such rugged densities are not only undesirable but they also may not relay the desired density structure. Several kernels which will be enumerated later in this study, have been constructed and applied to obtain smoothed curves. When a kernel is applied to the NDE, the kernel, signified by $K(\cdot)$ replaces the weight parameter $\omega(\cdot)$ in equation 5.5.

To describe a kernel, let x_1, x_2, \dots, x_n be IID random observations such that $x_i \in X \forall i = 1, 2, \dots, n$, having the pdf $f(x)$. The kernel estimate of $f(x)$ is defined

by:

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (5.6)$$

where K is the kernel and h is the bandwidth or the smoothing parameter. The kernel estimator is constructed by centering a scaled kernel at each observation, then the value of the kernel at each sample point x is given by the mean of the n kernel ordinates at the point x . The kernel therefore spreads out each peak by giving varying weights to all data points, proportional to their distance from the peak [71]. In this study we will only consider kernels which satisfy the following functional conditions [72]:

$$\int_{-\infty}^{\infty} K(u) du = 1 \quad \int_{-\infty}^{\infty} uK(u) du = 0 \quad \int_{-\infty}^{\infty} u^2 K(u) du \geq 0 \quad (5.7)$$

These kernel functions include [73]:

1. Box or uniform kernel: $K(u) = \frac{1}{2}I_{[-1,+1]}(u)$
2. Triangular kernel: $K(u) = (u + 1)I_{[-1,0]}(u) + (1 - u)I_{[0,+1]}(u)$
3. Quadratic or Epanechnikov kernel: $K(u) = \frac{3}{4}(1 - u^2)I_{[-1,+1]}(u)$
4. Biweight kernel: $K(u) = \frac{15}{16}(1 - u^2)^2I_{[-1,+1]}(u)$
5. Triweight kernel: $K(u) = \frac{35}{32}(1 - u^2)^3I_{[-1,+1]}(u)$
6. Gaussian kernel: $K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)$
7. Cosine kernel: $K(u) = \frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)I_{[-1,+1]}(u)$

In examining the list of the common kernels, it can be observed that all except the Gaussian kernel are bound between $[-1, 1]$. This implies that when estimating

the density function at a point (x) using the other kernels, appropriate data transformation is needed to ensure that:

$$\left| \frac{x - X_i}{h} \right| \leq 1 \quad (5.8)$$

The cumulative distribution function $F(x)$, is obtained from the probability distribution function $f(x)$, using the following relationship:

$$F(x) = \int_{\forall x \in X} f(x) dx$$

The cdf is then calculated from equation 5.6 as follows:

$$\hat{F}_n(x) = \frac{1}{nh} \sum_{i=1}^n \int_{-\infty}^{\infty} K\left(\frac{x - X_i}{h}\right) dx \quad (5.9)$$

It can be seen from equations 5.6 and 5.9 that the bandwidth, h controls the neighborhood that should be considered in the estimation procedure, and the choice of Kernel determines the performance of the kernel estimate. The next section discusses the rationale for the choice of the kernel function.

5.2 Choice of Kernel Estimate: Kernel Fitness Measure

The performance of a Kernel estimate requires the specification of some appropriate error criteria that globally measures how far the estimate is from the actual density function [74]. Rather than estimate the function f at single points, it is desirable to estimate f over its existing range. One such criterion is the Integrated

Squared Error (ISE) given by:

$$ISE[\hat{f}_n(x; h)] = \int_{-\infty}^{\infty} [\hat{f}_n(x; h) - f(x)]^2 dx \quad (5.10)$$

However, when the approximation is needed for several sets of data, it is appropriate to use the Mean Integrated Squared Error, MISE as follows:

$$MISE[\hat{f}_n(.; h)] = E[ISE[\hat{f}_n(.; h)]] = E \int_{-\infty}^{\infty} [\hat{f}_n(x; h) - f(x)]^2 dx \quad (5.11)$$

When the order of integration is changed, we get:

$$MISE[\hat{f}_n(.; h)] = \int_{-\infty}^{\infty} E[\hat{f}_n(x; h) - f(x)]^2 dx = \int_{-\infty}^{\infty} MSE[\hat{f}_n(x; h)] dx \quad (5.12)$$

5.3 Determination of Mean Square Error (MSE) of a Kernel Density Estimator

Let us replace $\hat{f}_n(.; h)$ with $\hat{\theta}$ and $f(x)$ with θ for the purpose of describing the Mean Square Error, and use $E(.)$ to signify the expected value. From statistical principles [75],

$$Variance(\hat{\theta}) = \frac{1}{n} \sum_{\forall i \in n} (\hat{\theta} - \theta)^2$$

which simplifies to:

$$Variance(\hat{\theta}) = E(\hat{\theta})^2 - 2\theta E(\hat{\theta}) + \theta^2$$

Also,

$$Bias^2(\hat{\theta}) = [E(\hat{\theta}) - \theta]^2$$

is simplified as:

$$Bias^2(\hat{\theta}) = [E(\hat{\theta})]^2 - 2\theta E(\hat{\theta}) + \theta^2$$

The MSE of an estimate $\hat{\theta}$ is given by:

$$MSE(\hat{\theta}) = E((\hat{\theta}) - \theta)^2$$

which simplifies to:

$$MSE(\hat{\theta}) = E(\hat{\theta})^2 - 2\theta E(\hat{\theta}) + E(\theta)^2$$

Since $E(\hat{\theta}) = \theta$ and $E(\theta) = E(\theta)^2 = 0$, the variance and the square of the bias sums up to:

$$Variance(\hat{\theta}) + Bias^2(\hat{\theta}) = E(\hat{\theta})^2 - 2\theta^2$$

Therefore,

$$MSE(\hat{\theta}) = Variance(\hat{\theta}) + Bias^2(\hat{\theta})$$

Likewise, using the density estimate notation $\hat{f}_n(.; h)$,

$$MSE[\hat{f}_n(x; h)] = Var[\hat{f}_n(x; h)] + Bias^2[\hat{f}_n(x; h)] \quad (5.13)$$

where

$$Bias[\hat{f}_n(x; h)] = E[\hat{f}_n(x; h)] - f(x) \quad (5.14)$$

5.3.1 Asymptotic MSE Approximation

The following assumptions are made to construct the large sample approximations of the variance and bias terms in the MSE as derived by Wand and Jones [72].

1. The density f is such that its second derivative f'' is continuous, integrable and ultimately monotone.

2. The bandwidth $h = h_n$ is a determined sequence of non-negative numbers, such that: $\lim_{n \rightarrow \infty} h = 0$ and $\lim_{n \rightarrow \infty} nh = \infty$.
3. The kernel is a bounded probability function whose fourth moment exists.

From equation 5.6, the $E\hat{f}(x, h)$ can be calculated as follows:

$$E[\hat{f}(x, h)] = E\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right] = E[K(x - X_i)] \quad (5.15)$$

Let x_i be replaced by y for ease of illustration, such that:

$$\begin{aligned} E[\hat{f}(x, h)] &= E[K(x - y)] \\ &= \int K_h(x - y) f(y) dy \end{aligned}$$

Let $(x - y)/h$ be z . Following this transformation:

$$E\hat{f}(x, h) = \int_{-\infty}^{\infty} K_h(z) f(x - zh) dz \quad (5.16)$$

According to Taylor's theorem which states that when $n \geq 0$ is an integer and f is a function which is n times differentiable on the closed interval $[a, x]$ and $n + 1$ times differentiable on the open interval (a, x) , the Taylor's series expansion of $f(x)$ is given by:

$$f(a - x) = \left\{ f(a) + f'(a)(x - a) + \frac{f''(a)(x - a)^2}{2!} + \dots + \frac{f^{(n)}(a)(x - a)^n}{n!} + R_n(x) \right\} \quad (5.17)$$

The last term is the error term which gets smaller at x nears a [72]. Using the following conditions,

$$\int_{-\infty}^{\infty} K(z) dz = 1, \quad \int_{-\infty}^{\infty} zK(z) dz = 0, \quad \text{and} \quad \int_{-\infty}^{\infty} z^2 K(z) dz < \infty \quad (5.18)$$

and the asymptotic assumptions in section 5.3.1, the Taylor's series expansion of the expectation term $f(x - zh)$ about x is:

$$f(x - zh) = \left\{ f(x) + f'(x)(-zh) + \frac{f''(x)(zh)^2}{2!} - \frac{f'''(x)(zh)^3}{3!} + o(h^2) \right\} \quad (5.19)$$

Substituting equation 5.19 in 5.16 we get:

$$E[\hat{f}_n(x; h)] = f(x) + \frac{1}{2}h^2 f''(x)\mu_2(k) + o(h^2) \quad (5.20)$$

Note that the Taylor expansion about x is terminated at the fourth moment following the asymptotic MSE convergence assumption number 3. Also, the second and the fourth terms disappear following the second condition in equation 5.18. The bias now becomes [76]:

$$Bias = \frac{1}{2}h^2 f''(x)\mu_2(k) + o(h^2) \quad (5.21)$$

where $a_n = o(b_n)$ means that $\lim_{n \rightarrow \infty} a_n$ is of order (b_n) iff $\lim_{n \rightarrow \infty} \left| \frac{a_n}{b_n} \right| = 0$, and μ_2 represents $\int_{-\infty}^{\infty} z^2 K(z) dz$ [77].

Similarly,

$$Var[\hat{f}_n(x; h)] = Var \left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right] \quad (5.22)$$

Using the variable transformation that was done for the expectation (see equation 5.16),

$$\begin{aligned}
Var[\hat{f}(x, h)] &= (nh)^{-1} \int_{-\infty}^{\infty} K(z)^2 f(x - zh) dz - n^{-1} [E\hat{f}(x, h)]^2 \\
&= (nh)^{-1} \int_{-\infty}^{\infty} K(z)^2 f(x) + o(1) dz - (n)^{-1} f(x) + o(1)^2 \\
&= (nh)^{-1} \int_{-\infty}^{\infty} K(z)^2 dz f(x) + o(nh)^{-1}
\end{aligned} \tag{5.23}$$

Therefore if we let $R(K) = \int_{-\infty}^{\infty} K(z)^2 dz$ we have:

$$Var[\hat{f}(x, h)] = (nh)^{-1} R(K) f(x) + o(nh)^{-1} \tag{5.24}$$

The bias, equation 5.21 is of order $o(h^2)$, which means that the function estimate is asymptotically unbiased whereas the variance, equation 5.24 is of order $(nh)^{-1}$ and therefore it converges to zero asymptotically [72]. From the bias 5.21 and variance 5.24 expressions we get:

$$MSE[\hat{f}_n(x; h)] = (nh)^{-1} R(K) f(x) + \frac{1}{4} h^4 \mu_2(K)^2 f''(x)^2 + o(nh)^{-1} + h^4 \tag{5.25}$$

Thus the Asymptotic Mean Integrated Square Error (AMISE) is given by:

$$AMISE[\hat{f}_n(., h)] = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} MSE[\hat{f}_n(x; h)] \tag{5.26}$$

$$= (nh)^{-1} R(k) + \frac{1}{4} h^4 \mu_2(k)^2 R(f'') \tag{5.27}$$

where $R(f'') = \int_{-\infty}^{\infty} f''(x)^2 dx$. The optimal choice of bandwidth h can be found by equating the partial derivative of AMISE with respect to h to zero. Thus,

$$h_{opt} = [\frac{R(k)}{\mu_2(k)^2 R(f'')n}]^{\frac{1}{5}} \quad (5.28)$$

and the corresponding AMISE is:

$$AMISE_{h_{opt}} = \frac{5}{4} [\sqrt{\mu_2(k)} R(k)]^{\frac{4}{5}} R(f'')^{\frac{1}{5}} n^{-\frac{4}{5}} \quad (5.29)$$

The optimal kernel function is found by minimizing the AMISE [69], [78]. Considering the common kernels, the Epanechnikov kernel is the most optimal and the performance (inefficiency) of the other common kernels relative to the Epanechnikov kernel is given in table 5.1 [72], [79].

Table 5.1. Relative inefficiencies of common kernels

Kernel	Inefficiency
Epanechnikov (quadratic)	1.000
Biweight	1.0061
Triangular	1.0135
Gaussian (normal)	1.0513
Box (Uniform)	1.0758

Since the inefficiencies are relatively similar, other factors such as the ease of computation and known properties of the estimate play significant roles in the choice of a kernel. In this work, we will use the Gaussian kernel based on the suggestion of Miladinovic and Tsokos. They proposed that the Gaussian Kernel is more robust as indicated by small changes in its AMISE value with increased sample size. See Miladinovic and Tsokos [80] for a detailed discussion of the stability of the Gaussian Kernel.

5.4 Bandwidth Selection

As it was mentioned in section 5.1, the choice of the smoothing parameter h dictates the degree of smoothing. With insufficient smoothing, meaning a narrow bandwidth h , the resulting density structure is exceedingly rough and contains spurious features that are perhaps artifacts of the sampling process. Conversely, with excessive smoothing, meaning a wide bandwidth, important features of the underlying structure are smoothed away [81]. From equations 5.21 and 5.24, we see that the choice of h introduces a tradeoff between the bias and the variance of the estimate. A smaller h results in a smaller bias and a larger variance and vice versa. Equation 5.28 indicates that the optimal value of h depends on the underlying distribution, which is often unknown. Therefore, in this study we explore data-specific optimum choice of h .

Several approaches to bandwidth selection have been explored over the years. They are broadly categorized as: (i) first-generation approaches (proposed before 1990), which include the rule of thumb method [71], least squares cross-validation [82], and biased cross-validation [83]. (ii) second-generation (proposed after 1990) which exhibit superior performance than the first generation, and consists of the solve-the-equation plug-in approach [84] and smooth bootstrap method [85]. Marron et al. compared the performance of the two categories of bandwidth choice methodologies using three validation procedures, namely using real data, simulated data, and asymptotic analysis [81]. Their research indicated that the second category of methods outperformed the first over the three validation procedures. Silverman [71] proposed a constant optimal bandwidth by replacing the unknown term, $R(x)$ in h_{opt} by an estimate based on the kernel function used (see equation 5.28). For instance, for the Gaussian kernel,

$$\hat{h}_0 = 1.06 \times \min \left(\hat{\sigma}, \frac{\hat{R}}{1.34} n^{0.5} \right)$$

where $\hat{\sigma}$ is the sample standard deviation and \hat{R} is the sample interquartile range.

Liu et al. [69] applied Silverman's idea to have a natural bandwidth candidate h_1 , then visually proposed an arbitrary integer l which is the number of data partitions. They then created a vector of band widths h_i such that $h_i = \frac{ih_1}{l}$ for $i = 1, 2, \dots, l$. All the bandwidths were tried on the Gaussian kernel and the optimal one was chosen using visual judgment.

Shimazaki and Shinomoto [86] suggested a method for selecting a constant optimal bandwidth by minimizing the MISE between the estimated and the underlying rates of action potentials in brain activity. In this paper, we extend their approach to develop a variable bandwidth using the Gaussian kernel so as to estimate the underlying pdf using univariate dielectric failure data.

5.5 Variable Optimal Bandwidth Selection

Variable bandwidths are a good alternative for density estimation especially when differential spatial smoothing is desired. Since the bias is directly proportional to the bandwidth, varied smoothing will allow for possible reduction of bias in the highly peaked regions and a corresponding reduction in variance in the flat regions. This flexibility enables local fitting of polynomials so that they can adapt to spatially non homogeneous curves [78].

We propose to determine the variable bandwidth by minimizing a cost function which is derived from the MISE. As shown earlier, the MISE is expressed as:

$$MISE[\hat{f}_n(., h)] = \int_{-\infty}^{\infty} E[\hat{f}_n(x; h) - f(x)]^2 dx = \int_{-\infty}^{\infty} MSE[\hat{f}_n(x; h)] dx \quad (5.30)$$

The integrand in equation 5.30 can be decomposed into three terms as follows:

$$MISE[\hat{f}_n(., h)] = \int_{-\infty}^{\infty} E[\hat{f}_n(x; h)]^2 - 2E[\hat{f}_n(x; h)f(x)] + [f(x)]^2 dx \quad (5.31)$$

Since the last term of the integrand does not depend on the chosen kernel, we can eliminate it from the MISE expression and set the cost function, $C_n(h)$ to:

$$C_n(h) = \int_{-\infty}^{\infty} E[\hat{f}_n(x; h)]^2 dx - 2 \int_{-\infty}^{\infty} f(x)E[\hat{f}_n(x; h)] dx \quad (5.32)$$

Bowman [82] shows that the Least Square Cross-Validation (LSCV) estimate of MISE is given by:

$$LSCV(h) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} [\hat{f}_n(x; h)]^2 dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{n,-i}(X_i; h) + 2n^{-1} \sum_{i=1}^n f(x_i) \quad (5.33)$$

where

$$\hat{f}_{n,-i}(X_i; h) = \frac{1}{n-1} \sum_{j \neq i}^n K\left(\frac{x - X_i}{h}\right)$$

Therefore, removing the last term from the LSCV equation and replacing $\hat{f}_n(x; h)$ with the kernel function, the cost function becomes:

$$C_n(h) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} [k(x - x_i)k(x - x_j) dx - 2n^{-1} \sum_{i \neq j} k(x_i - x_j)] \quad (5.34)$$

Any symmetrical kernel function including the Gaussian kernel is invariant to the exchange of x_i and x_j when computing $k_h(x_i - x_j)$. In addition, the covariance in equation 5.34 is symmetrical with respect to i and j . Thus we obtain:

$$\sum_{i \neq j} k_h(x_i - x_j) = 2 \sum_{i < j} k_h(x_i - x_j) \quad (5.35)$$

Substituting equation 5.35 in 5.34 we get:

$$C_n(h) = n^{-1} \sum_{i=1}^n \int_{-\infty}^{\infty} [k(x - x_i)k(x - x_j) dx - 4n^{-1} \sum_{i < j} k(x_i - x_j)] \quad (5.36)$$

In order to remove the necessity for integration in equation 5.36, the Gaussian kernel is denoted as:

$$N(x, h^2) = (2\pi)^{-\frac{1}{2}} h^{-1} \exp \frac{-x^2}{2h^2} \quad (5.37)$$

and whose integration can be carried out analytically is used. This reduces equation 5.36 to

$$C_n(h) = \frac{1}{n-1} N(0, h^2) + \frac{n-2}{n(n-2)^2} \sum_{i < j} N(x_i - x_j, 2h^2) - \frac{2}{n(n-1)} \sum_{i < j} N(x_i - x_j, h^2) \quad (5.38)$$

Equation 5.38 is further simplified and the resultant cost function becomes:

$$2\sqrt{\pi}n^2\hat{C}_n(h) = \frac{n}{h} + \frac{2}{h} \sum_{i < j} \exp \frac{-(x_i - x_j)^2}{4h^2} - 2 \exp \frac{-(x_i - x_j)^2}{2h^2} \quad (5.39)$$

5.5.1 Optimum Bandwidth Selection Methodology

In this section, we will present two methods, namely: (i) the procedure to obtain a constant bandwidth, proposed by Shimazaki et al. [86], and (ii) our procedure for obtaining a variable bandwidth, as shown in figure 5.1.

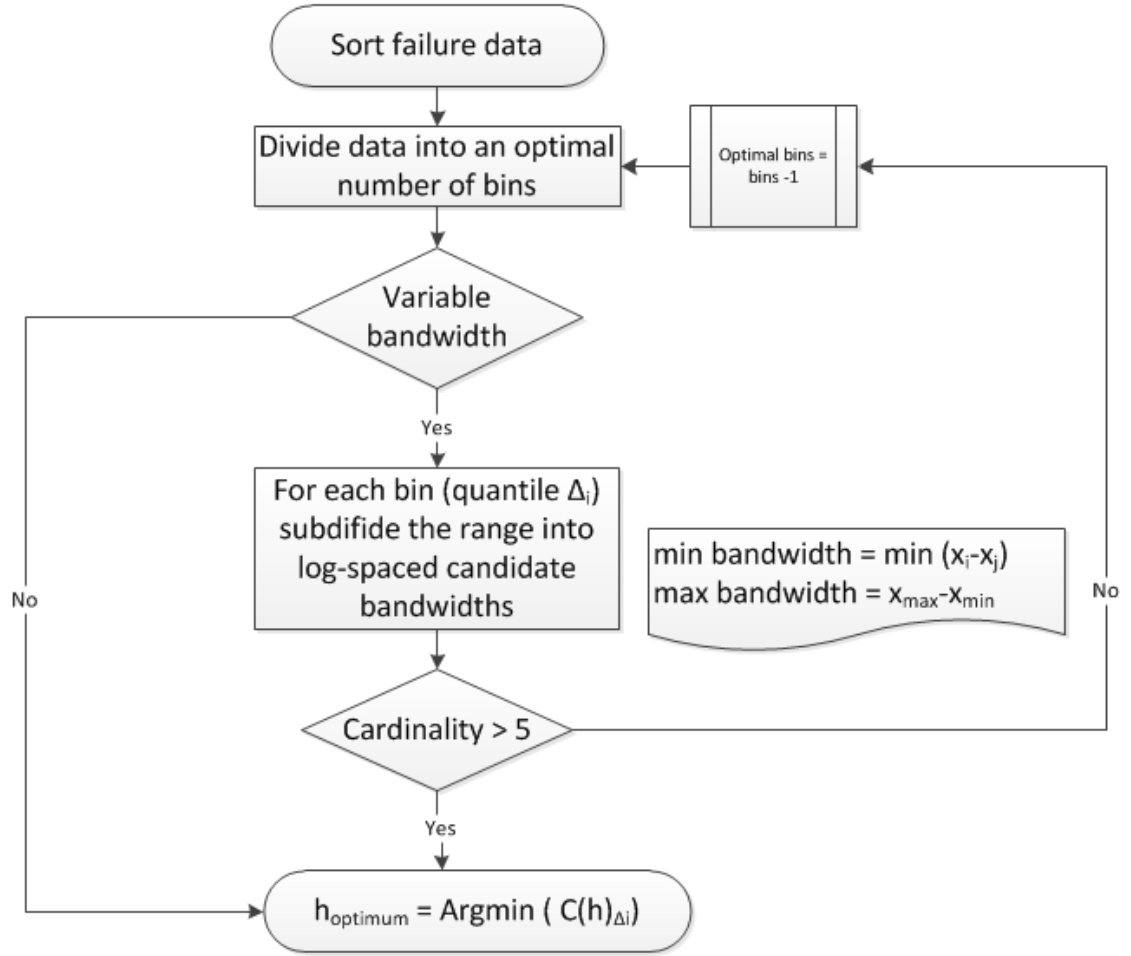


Figure 5.1. Procedure for bandwidth optimization

5.5.1.1 Constant Bin Width Selection Procedure

The following procedure was adopted from Shimazaki et al. [86] to determine the choice of an optimal constant bandwidth.

1. Arrange the failure data in an ascending order $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.
2. Develop log-spaced default intervals h_j ranging from $\min(x_{(i+1)} - x_{(i)})$ for $i = 1, 2, \dots, n$, to the entire data range $(x_{\max} - x_{\min})$. Where $j = 1, 2, \dots, H$.

3. For each interval, calculate the corresponding cost function $C(h_j)$ and choose the optimal bandwidth, $h_{opt} = \operatorname{argmin} C(h)$.
4. Fit a KDE using the constant h_{opt} .

5.5.1.2 Variable Bandwidth Selection Procedure

We propose the following procedure to determine the choice of a variable bandwidth.

1. Arrange the failure data in an ascending order $x_{(1)}, x_{(2)}, \dots, x_{(n)}$.
2. Find the optimal number of bins H for the usual data histogram using Scott's choice [87] or Freedman-Diaconis' choice criteria [88].
3. Divide the failure data or its log transform into H equispaced quantiles Δ_j such that each quantile contains at least two data points, where $j = 1, 2, \dots, H$.
4. For each quantile do the following: Define the initial default bandwidth vector for each interval by developing log-spaced bandwidths h_k ranging from $\min(x_{(i+1)} - x_i)$ to the entire data range $x_{\max} - x_{\min}$ for $\forall x \in \Delta_j$.
5. For each bandwidth size in the default vector, calculate the corresponding cost function and choose the variable bandwidth as $h_{opt, \Delta_i} = \operatorname{argmin}(C(h)_{\Delta_i})$.
6. Fit a bounded KDE to determine the estimated density function in each interval $\hat{f}x, \Delta_i$.
7. Augment the $\hat{f}x, \Delta_i$ values over all Δ_i intervals, and plot the resultant pdf against the failure times.

5.6 Numerical Results

We used high-k/Si stack MOSCAP device failure data obtained from Luo et al. [6] to construct KDEs using Gaussian kernel, and compared the resultant KDEs using constant bandwidths versus variable bandwidths for three univariate failure data obtained at 8.1, 7.9 and 7.7 MV/cm electric field levels. The results of the constant and variable bandwidths for the dielectric failure data at each stress level are summarized in table 5.2. Figure 5.2 shows the difference in the estimated density structure when the constant and variable bandwidth were used.

Table 5.2. KDE constant and variable bandwidth for failure data at different stress levels

Stress level	Constant Bandwidth	Variable Piece-wise Constant Bandwidth			
	Bandwidth (sec)	1 st Quartile	2 nd Quartile	3 rd Quartile	4 th Quartile
8.1 MV/cm	80	8.7	99.7	343.8	562.8
7.9 MV/cm	123	66.4	247.0	737.0	749.6
7.7 MV/cm	298	152.1	184.6	2652.0	2074.0

It can be seen from figure 5.2 that the variable bandwidth is very sensitive to peaks within the failure data, because it tends to select narrow optimal bandwidths in sections of the data with most points, and wide optimal bandwidths in areas with sparse points. This result is also associated with the multiple peaks on the left side of the density estimate plots of figure 5.2. However, despite peak sensitivity, the general form of the functions from the constant bandwidth matched the variable bandwidth kernel estimates.

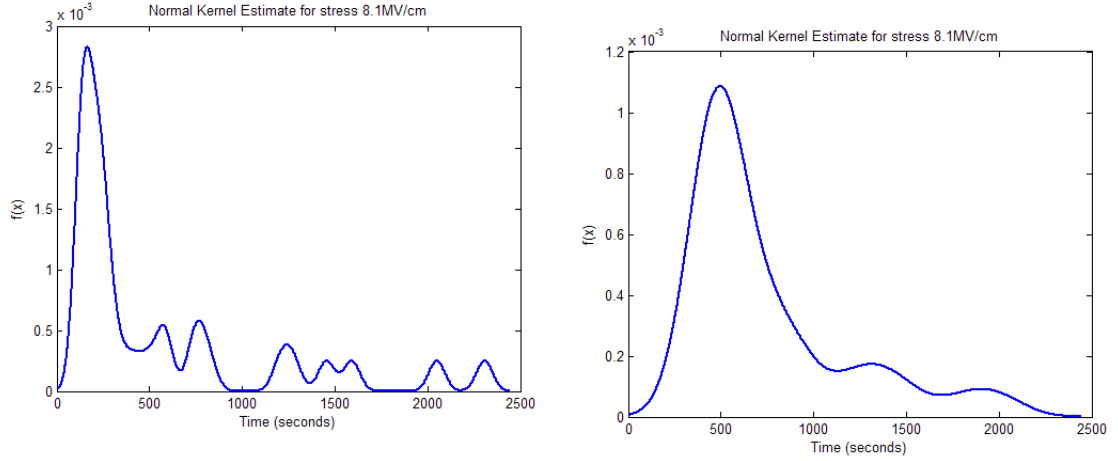


Figure 5.2. KDE with constant (right) and variable (left) bandwidth for failure data at 8.1 MV/cm electric field stress level

5.7 Reliability Inference and Numerical Results

The reliability function $\hat{R}_n(x)$ is calculated from the cumulative failure density function $\hat{F}_n(x)$ from equation 5.9 as follows:

$$\hat{R}_n(x) = 1 - \hat{F}_n(x) \quad (5.40)$$

Therefore the reliability estimate is defined by:

$$\hat{R}_n(x) = 1 - \frac{1}{nh^*} \sum_{i=1}^n \int_{-\infty}^x K\left(\frac{x - X_i}{h^*}\right) dx \quad (5.41)$$

where h^* is the optimal variable bandwidth obtained in table 5.2.

In this study, we estimated cumulative density function $\hat{F}(t)$ and the reliability function $\hat{R}(t)$ using the variable optimal bandwidths for data at each stress level, as illustrated in figure 5.3. As expected, figures 5.3 and all the other reliability and cdf density estimates included in Appendix A are typical reliability and cdf functions

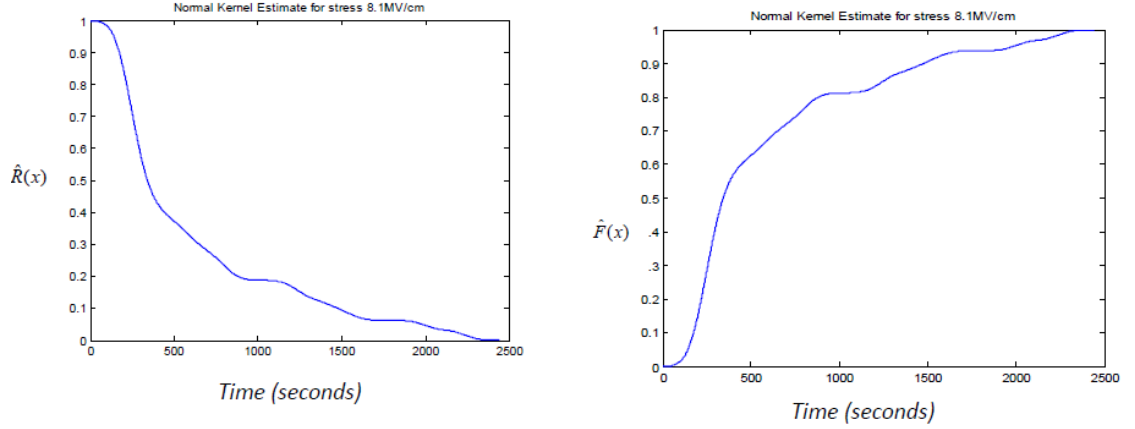


Figure 5.3. Reliability and cdf estimates at 8.1 MV/cm electric field stress level

which indicate an initial highly increasing failure probability followed by near-constant failure probability.

5.8 Conclusion

Kernel density estimation methods are useful for reliability inference, especially for failure phenomena whose data do not conform to traditional probability distributions. In this work, we have discussed nonparametric function estimation using the kernel approach. We have shown that based on the criteria of minimizing the Mean Integrated Squared Error (MISE), the efficiencies of the commonly used symmetric kernels are relatively the same. For this reason, the most important performance measure of kernel estimates is the choice of the bandwidth.

We develop a cost minimization algorithm for selecting a vector of optimal variable bandwidths, optimized at each failure data sub-interval. In addition, we compare the performance of (fixed) constant bandwidth versus variable bandwidth in estimating failure density. Unlike fixed bandwidths, variable bandwidths are flexible and are optimized according to the data in the following way: each data point within a

sub-interval contributes to the optimal bandwidth within that interval, and these bandwidths are variably selected for each sub-interval.

We believe that this approach renders the resultant density structure fit for the failure data. We tested the variable bandwidth selection algorithm against accelerated failure data for different CMOS devices: (i) that have a high permittivity (high-k) dielectric film and (ii) silicon carbide thin films. We believe that the need for reliability inference for nanomaterials is a pertinent research field especially with the current aggressive development of newer nanomaterials and nanodevices progresses. This study provides a framework for constructing realistic failure characterization and nanoreliability inference tools.

CHAPTER 6

BAYESIAN INFERENCE

6.1 Background

Unlike classical statistical approaches that make inference from observations alone, Bayesian approaches provide enrichment to statistical inference by formalizing the process of learning from historical information to update the most optimal estimates of the parameters of a stochastic process [89], such as the location, shape and the scale parameters. This is made possible by the provision of a framework to sequentially update the parameter estimates given extra acquired information [90]. Common classical inference techniques include maximum likelihood, least square error method, and method of moments.

The Bayesian methodology starts with a prior distribution, $G(\theta)$ of the unknown random variable θ in the sample space Θ . Let X be a random variable that has a probability structure that depends on θ . Let X_1, \dots, X_n denote a random sample from the distribution of X and let T denote a statistic which is a function of X_1, \dots, X_n . Then the conditional probability density function of T for every $\theta \in \Theta$ is given by $f(t|\theta)$ and is known as the likelihood function. The conditional pdf of θ given T is the posterior distribution is given by:

$$p(\theta|T) = \frac{f(t|\theta)G(\theta)}{\int_{\Theta} f(t|\theta)G(\theta) d\theta} \quad (6.1)$$

6.2 Choice of Likelihood Function

Given the high-k dielectric failure data, goodness-of-fit tests were performed to determine the most appropriate likelihood function. Given that the random variable of interest is time to failure, which is nonnegative and continuous, candidate likelihood functions for failure data must be defined in the $[0, \infty]$ range [44]. With this in mind, Weibull, log-normal, gamma, extreme value and normal distributions were fitted to the data using the maximum likelihood estimation (MLE) methodology. The MLE approach determines the unbiased and efficient estimates of the given set of parameter Θ , by maximizing the likelihood function $L(\Theta)$, or its logarithm, $\ln(L(\Theta))$. For observed data t , the likelihood function, $L(\Theta) = f(t|\Theta)$, considered as a function of Θ [91] [92]. For the MLE method, the decision criterion is to choose the distribution that results in the highest log-likelihood fitness measure. Table 6.1 shows the summary of the log-likelihood values.

Table 6.1. Log-likelihood values for different distributions at given stress levels

Distributions	8.1 MV/cm	7.9 MV/cm	7.7 MV/cm	7.5 MV/cm	7.3 MV/cm	7.1 MV/cm
Weibull	-222.28	-231.79	-281.75	-305.69	-303.43	-323.33
Log-normal	-222.96	-233.66	-283.42	-308.53	-305.23	-323.70
Gamma	-222.49	-231.54	-281.84	-305.61	-303.69	-323.77
Extreme value	-261.29	-265.61	-319.67	-329.59	-348.63	-365.96
Normal	-252.84	-259.83	-310.67	-321.98	-335.97	-354.33

From table 6.1, it is seen that the Weibull, log-normal and gamma distributions have higher log-likelihoods than the extreme value and normal distributions for all stress levels. More goodness-of-fit analyses were carried out to provide better discrimination among the Weibull, log-normal and gamma distributions using the Kolmogorov-Smirnov (K-S) tests. The K-S test is a distribution-free curve fitting approach that applies to continuous distributions, and has been shown to provide superior estimates of error in curve fitting models [44]. By using the K-S test, we were able to

demonstrate marked differences in the fitness of the distributions. Table 6.2 contains a summary of the resulting p-values and the K-S statistic corresponding to each of the distribution. These results show that the Weibull distribution has the higher p-value, an indication of its fitness superiority over the log-normal and the gamma distribution.

Table 6.2. Results of K-S goodness-of-fit tests for Weibull, log-normal and gamma distributions

	Electric field (in MV/cm)					
Weibull fit	8.1	7.9	7.7	7.5	7.3	7.1
Scale parameter	336.57	741.531	1301.709	2807.52	2451.662	3354.396
Shape Parameter	0.615	0.583	0.667	0.87	0.654	0.683
Log-likelihood	-222.287	-231.795	-281.759	-305.692	-303.436	-323.336
K-S test statistics	0.212	0.166	0.176	0.117	0.117	0.114
K-S test p-value	0.569	0.76	0.621	0.962	0.962	0.967
Log-normal fit						
Location parameter	4.889	5.597	6.303	7.262	6.925	7.294
Shape Parameter	1.963	2.201	1.874	1.504	1.91	1.732
Log-likelihood	-222.962	-233.665	-283.429	-308.538	-305.23	-323.706
K-S test statistics	0.187	0.2	0.147	0.147	0.176	0.2
K-S test p-value	0.58	0.537	0.825	0.825	0.621	0.441
Gamma fit						
Scale parameter	977.293	2418.089	3115.647	3765.605	6195.653	7630.584
Shape Parameter	0.493	0.456	0.548	0.797	0.533	0.572
Log-likelihood	-222.495	-231.544	-281.849	-305.613	-303.693	-323.775
K-S test statistics	0.225	0.2	0.205	0.147	0.205	0.185
K-S test p-value	0.35	0.537	0.422	0.825	0.422	0.599

Based on the goodness-of-fit tests results, we will construct the Bayesian inference framework using the Weibull distribution function. However, before presenting this framework, we describe the physical meaning of the properties of the Weibull distribution in the context of dielectric breakdown.

6.3 Two-Parameter Weibull Distribution in the Context of Dielectric Breakdown

Defect-based dielectric failure is triggered when stress-induced defects are introduced into the dielectric. The reason for using the Weibull distribution follows from the assumptions that: (i) failure occurs whenever a critical amount of defects is reached, enough to form conductive paths that bridge the dielectric bulk, and (ii) the defects are uniformly distributed throughout the dielectric layer [61].

The 2-parameter Weibull distribution is expressed as follows [93]:

$$f(t_i|\theta, \beta) = \frac{\beta}{\theta} \left(\frac{t_i}{\theta}\right)^{\beta-1} \exp - \left(\frac{t_i}{\theta}\right)^{\beta} \quad (6.2)$$

where $\theta > 0$ is the scale parameter, $\beta > 0$ is the shape parameter and $t_i > 0 \forall i = 1, 2, \dots, n$, are the failure times. In reference to Bayesian analysis, the Weibull failure distribution in equation 6.2 is used to develop the likelihood function given by:

$$L(\theta, \beta) = \left(\frac{\beta}{\theta}\right)^n \prod_{i=1}^n \left(\frac{t_i}{\theta}\right)^{\beta-1} \exp \left(-\frac{\sum_{i=1}^n t_i}{\theta}\right)^{\beta} \quad (6.3)$$

Henceforth, we will refer to the 2-parameter Weibull distribution as the Weibull distribution. In equation 6.2, the Weibull distribution is characterized by the scale parameter, θ and the shape parameter, β .

6.3.1 Weibull Shape Parameter

The shape parameter β , is defined in the $(0, \infty)$ range and it is dimensionless. With reference to dielectric failure process, we assume the shape parameter β is

proportional to the dielectric thickness, t_{diel} as expressed by:

$$\beta = \varsigma \left(\frac{t_{diel}}{a_0} \right) \quad (6.4)$$

where a_0 is the simulation cubic lattice constant, equal to the defect diameter and the proportionality constant, ς , is assumed to be approximately 0.38 [63], [62]. Table 6.3, presents a summary of the maximum likelihood β estimates of the simulated failure times for a given dielectric thickness. These results are in agreement with equation 6.4, and they show that β increases with an increase in dielectric thickness. Physically, this means that the dielectric failure rate increases with time, and the higher the value of β the faster the rate of failure increases with time.

Table 6.3. Shape parameter estimates of simulated failure time data

Simulated t_{diel}	1 nm	2 nm	3 nm	4 nm	5 nm
β	0.85	1.201	1.62	2.08	2.50

6.3.2 Weibull Scale Parameter

The scale parameter θ consists of real values in the $(0, \infty)$ range and is the time at which the probability of failure for the device is 63.2% [93]. For this reason it is referred to as the product characteristic life. In the next section, we will discuss the physical failure model using the Arrhenius degradation model proposed by Nelson [34] and its relationship to the Weibull scale parameter θ .

6.3.2.1 Arrhenius-Weibull Model

The Arrhenius-Weibull model is a physical-statistical model that combines the Weibull life description with the Arrhenius dependence of dielectric life on the stress

conditions, which, in our case, is the amount of electric field applied to the dielectric E_l [34]. The use of this model requires the following assumptions: (i) at any stress level, the product life, indicated by the failure times, has a Weibull distribution and equivalently, the natural logarithm of the failure times follows the extreme value distribution, (ii) the Weibull shape parameter is constant at all stress levels, and (iii) the natural logarithm of the dielectric's characteristic life is a linear function of the stress level [34].

Based on the Arrhenius failure acceleration model, the high-k dielectric breakdown is assumed to be a thermo-chemical process, such that the rate of dielectric failure is expressed as a function of the activation energy or the enthalpy of activation ΔH (measured in eV), test temperature, T (measured in Kelvin, K), Boltzmann constant, k (measured in eV per Kelvin, i.e., eV/K), applied electric field E_l (measured in MV/cm), and the field acceleration parameter γ (measured in cm/MV) [52] as follows:

$$\tau_{bd} \propto \exp - \left[\frac{\Delta H}{kT} - \gamma E_l \right] \quad (6.5)$$

Nelson et al. indicate that the motivation for the Arrhenius rate law stems from the fact that failure occurs after a critical amount of reaction or degradation [34]. Therefore,

$$(\text{critical amount of degradation}) = (\text{rate of degradation}) \times (\text{time of failure})$$

and equivalently,

$$(\text{time of failure}) = (\text{critical amount of degradation}) / (\text{rate of degradation})$$

We will assume that the dielectric degradation process meets the requirements of the Arrhenius rate law, in which case we will assume that the life of the dielectric, t_{bd} is

inversely proportional to the reaction rate, τ_{bd} in equation 6.5 as follows [5], [34]:

$$t_{bd} \propto \exp \left[\frac{\Delta H}{kT} - \gamma E_l \right] \quad (6.6)$$

If the first term in the exponent is replaced by a constant, (it is dimensionless and hence a constant) and assuming that the estimation error is multiplicative, then we can linearize equation 6.6 as follows:

$$\ln(t_{bd}) = \alpha - \gamma E_l + \delta_l \quad (6.7)$$

where δ_l is the estimation error. Since the life characteristic θ (63.2th) percentile, represents a specific time on the time axis, it is reasonable to assume that it follows the same relationship as time in equation 6.7, thus:

$$\ln(\theta) = \alpha - \gamma E_l + \delta_l \quad (6.8)$$

Equation 6.13 will be useful later in the chapter, for developing the Bayesian model.

6.3.3 Weibull Plot

A quick and simple way to test if the unknown distribution underlying failure data is in fact the Weibull distribution is by constructing the Weibull plot, which is a graph of $(\frac{1}{1-\hat{F}(t_{(i)})})$ versus time to failure on a log-log scale paper [48]. Here, $\hat{F}(t_{(i)})$ is the empirical cumulative distribution function (ECDF) estimate determined using the median-rank statistics. If the dielectric failure distribution is indeed Weibull, the resultant Weibull plot should be a straight line, whose slope is equivalent to the Weibull shape parameter, and the dielectric characteristic life can be interpolated from the graph as the 63.2% failure time.

Figure 6.1, shows the Weibull plots of the simulated failure data for 1, 2, 3, 4 & 5 nm thickness. The dielectric characteristic time is determined by drawing a horizontal line from the 63.2 percentile mark (on the y-axis), to meet the population line, then drawing a vertical line from this intersection to the x-axis. The points at which the vertical line cuts the x-axis is the approximate dielectric characteristic life for a given thickness.

Figure 6.1 indicates that the dielectric characteristic life θ , increases with dielectric thickness. This is because for a constant stress level, thinner dielectric films will tend to breakdown faster than thicker films. Also, the figure shows that the slope, which is the Weibull shape parameter, β , increases with increasing dielectric thickness as we proposed in equation 6.4. Overall, figure 6.1 shows that it takes longer to realize the characteristic life for thick dielectrics. However, thick films exhibit increasing failure rates than thin films.

The performance of the simulation is sensitive to the number of defects that are introduced in the model at each iteration. In our simulation, we initially released a random number of defects following a Poisson distribution, with a mean of 100 defects at each iteration, for all the simulated thickness. The resultant failure times from this initial simulation with 1000 replicates for each thickness, (each replicate symbolizing a dielectric sample), produced the Weibull plot in figure 6.2. The population lines that correspond to 1, 2, and 3 nm thickness are not straight, and the misalignment, is attributed to the introduction of a large number of defects (a mean of 100 at each iteration). Given this problem we reduced the mean number of defect to 10 at each iteration, in the subsequent simulations. This resulted in a better Weibull plot as shown in figure 6.1.

Figure 6.3 is the Weibull plot of real dielectric failure data obtained from Luo et al. [6]. They developed Al/high-k/Si capacitor structures using 60 nm Hf-based dielectric

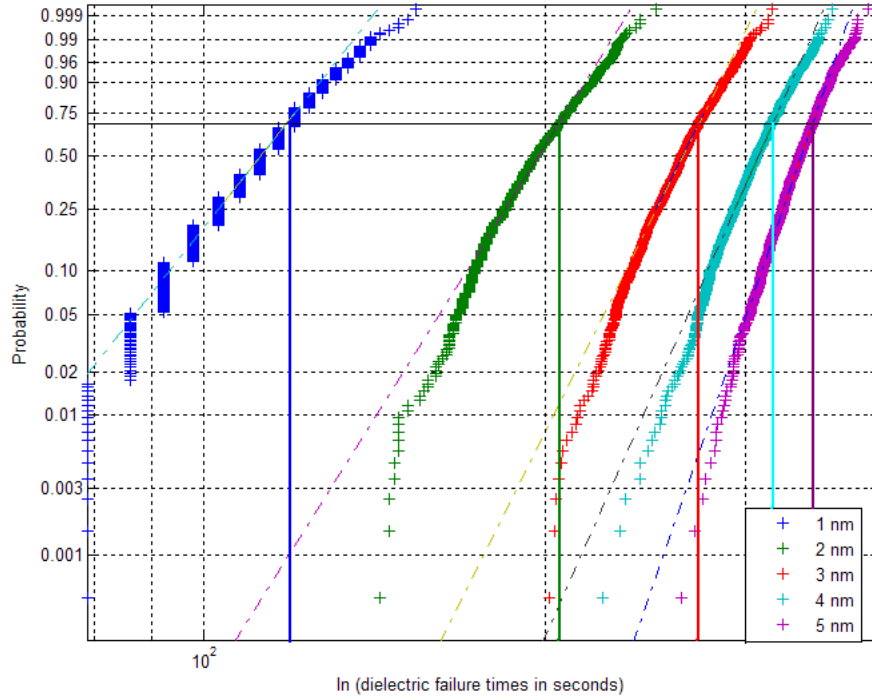


Figure 6.1. Weibull plot of simulated failure data for 1, 2, 3, 4 & 5 nm dielectric thickness with 26 degrees of communication

layers (this translates to approximately 2.19 nm SiO_2 equivalent oxide thickness). These samples were stressed at 8.1, 7.9, 7.7, 7.5, 7.3 and 7.1 MV/cm electric field levels.

Figure 6.3 is in agreement with the Arrhenius-Weibull model, because increased stress levels decrease the dielectric failure times and hence the dielectric characteristic life. Ideally, the slopes of these plots should be equal, signifying that the shape parameter is constant at all stress levels, and that increased stresses do not induce different failure modes. However, most experts argue that in reality, failure modes and mechanism get altered at higher stress levels [34]. For instance, while the Weibull plots in figure 6.3, that correspond to the data acquired at 8.1, 7.9 and 7.3 MV/cm tend to be parallel, those that correspond to data from 7.7, 7.5 and 7.1 MV/cm intersect, indicating that either a different failure mode was triggered at these stress

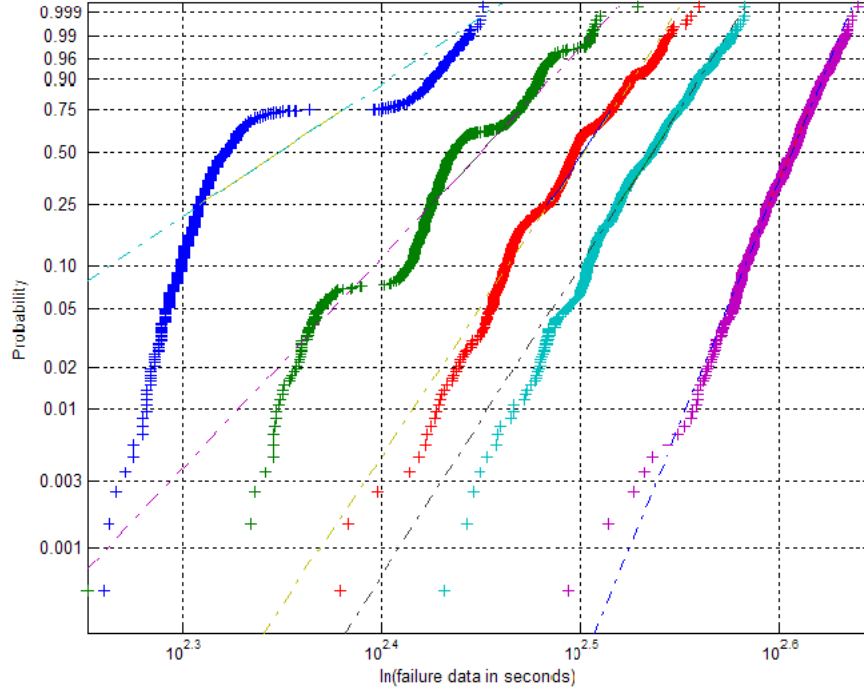


Figure 6.2. Weibull plot of simulated failure data for 1, 2, 3, 4 & 5 nm dielectric thickness with increased defect generation

levels, or there were erroneous data points following variation in the fabrication and testing processes. In this study, we will assume, that the shape parameter β , remains constant at all stress levels, and that increasing the test stress level at constant dielectric thickness, does not significantly alter the dielectric failure mode [34]

Although the Weibull parameters the shape, β and characteristic life, θ can be estimated directly from the Weibull plots, these estimates depend on graphical interpolation and are therefore subjective. In addition, Weibull plots often require large data sets to make meaningful statistical inference [94]. However, most life tests have small data sets due to the small number of failures during the specified test cycle. Given this challenge, the Bayesian approach offers an alternative solution.

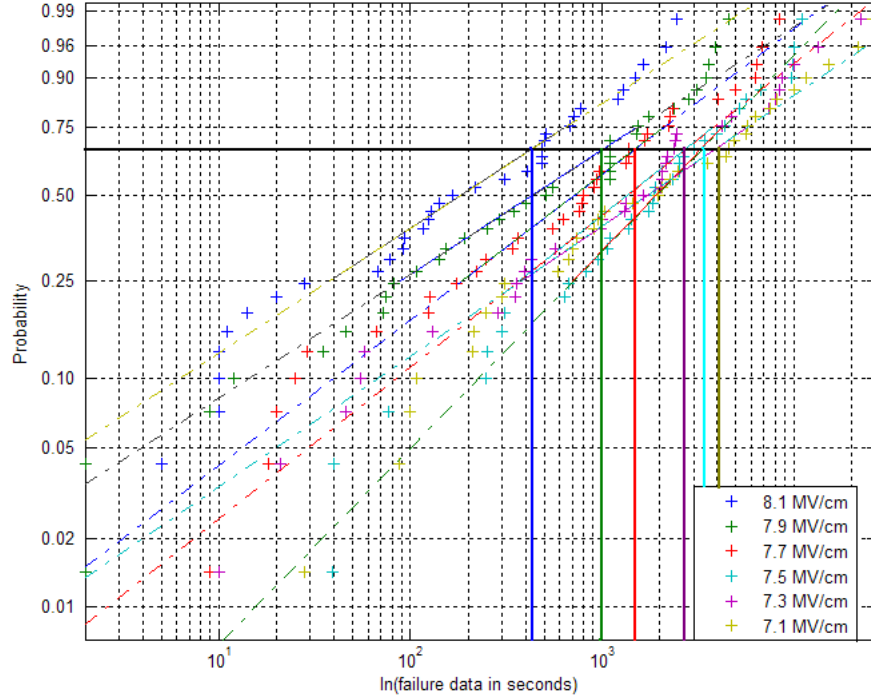


Figure 6.3. Weibull plot of actual dielectric failure data from accelerated degradation tests

6.4 Choice of Prior Distributions

One of the predicaments in implementing the Bayesian framework is the identification, selection and justification of a prior [95]. Savchuk and Tsokos summarize the criteria for the choice of prior to be based on two premises namely the information criterion, and noninformative criterion [96]. Typically, informative priors are specified based on historical information regarding the underlying stochastic process or from expert opinion. Depending on the nature of the prior parameters, whether discrete or continuous, an assumption is made about the nature of the family of known distributions to which the prior belongs. Such priors are referred to as informative priors. A special case occurs when the class of distribution is strategically chosen, so that the convolution of the prior distribution with the likelihood function results in a

posterior distribution from the same family of distribution as the prior. Such priors are said to be natural conjugates and often result in posterior functions that either have known probability structures or whose form can easily be estimated [92]. The drawback of using such priors is that there might not be a valid class of distribution functions suitable for a given set of data.

Noninformative priors provide alternative priors in cases where there is no a priori information regarding the unknown parameters or when there is no valid candidate structures for the prior distributions [90]. These vague priors, as their names indicate, give equal weights to all admissible values of the unknown parameters. For instance, the noninformative prior may be bounded, such as $\pi(\theta) \sim \text{uniform}(a, b)$, or unbounded, such as $\pi(\theta) \sim \text{uniform}(-\infty, \infty)$ and $\pi(\ln\theta) \sim \text{uniform}(0, \infty)$. While the uniform prior is possibly the simplest form of a vague prior that represents the ignorance regarding the values and variability of the unknown parameter, Carlin et al. note that the uniform prior is not invariant to transformations that occur in re-parameterized models [97].

A special kind of a noninformative prior is the Jeffrey's prior defined by [98]:

$$\pi(\theta) \propto [I(\theta)]^{\frac{1}{2}} \quad (6.9)$$

where $I(\theta)$ is the expected Fisher information of the probability distribution function, given by:

$$I(\theta) = -E_{t|\theta} \left[\frac{\partial^2}{\partial \theta^2} \log f(t|\theta) \right] \quad (6.10)$$

In the case of Jeffrey's prior, the choice of the sample determines the choice of prior through the expected value $E_{t|\theta}$. While the Jeffrey's prior is invariant to model re-parameterizations and variable transformations, it induces posterior computational complexities [92].

In this study, the proposed Bayesian model utilizes the two-parameter Weibull likelihood function with noninformative priors. When the Weibull parameters, θ and β are both assumed unknown, a prior should be placed on (θ, β) . As mentioned earlier, it is generally desirable that the joint prior should belong to a family of distributions that has a closed form, such that if the prior is conjugate, the resultant posterior will have a tractable form [99]. A number of experts argue that it is extremely difficult to find a family of continuous joint prior distributions consisting of both parameters, that is closed under sampling, in order to ensure that the ensuing convolution results in a legitimate posterior distribution [95], [100]. Given this difficulty, we propose the bounded uniform prior, given that there is limited information regarding the underlying structure of the unknown parameters θ , the characteristic life, and β , the Weibull shape parameter. As such, there is little justification to choose a specific prior with known distribution. In our case, the only subjective information that is possible is perhaps the range of values that the parameters can assume.

Normally in Bayesian inference problems, the specification of the prior is important since it influences the posterior by providing the initial parameter estimates upon which the likelihood is conditioned [101]. The parameters of the likelihood function are determined by the probability structure of the prior distribution. Specifically, when there is strong a priori belief regarding the values of elicited parameter estimates, then the chosen prior distribution strongly influences the likelihood function. However, in case of considerable uncertainty regarding the prior parameter estimates such as in our situation, a vague prior, with minimal influence on the likelihood and eventually the posterior distributions is used. We therefore opted to use the bounded uniform prior, which gives equal probabilities to all the parameter values within the admissible parameter space.

From equation 6.11, the posterior distribution when the bounded uniform prior is used is as follows:

$$p(\Theta|t) = \frac{f(t|\Theta)G(\Theta)}{\int_{\forall\Theta} f(t|\Theta)G(\Theta) d\Theta} \quad (6.11)$$

where $G(\Theta)$ is a bivariate uniform prior distribution and Θ is a vector of θ and β .

6.5 Hierarchical Bayesian Implementation

In this study, we will develop a hierarchical Bayesian model, which provides a way to incorporate subjective structural information about the unknown parameters in the likelihood function, in our case, the characteristic life, θ and the shape parameter, β .

In equation 6.13, the Arrhenius-Weibull model is used to describe the relationship between the dielectric characteristics life θ and the stress level, E as follows:

$$\ln(\theta) = \alpha - \gamma E + \delta \quad (6.12)$$

where $\alpha = \frac{\Delta H}{kT}$, and ΔH is the activation energy or the enthalpy of activation (measured in eV), T is the test temperature, (measured in Kelvin, K), and k is the Boltzmann constant, (measured in eV per Kelvin, i.e., eV/K).

In the hierarchical Bayesian framework, α , γ , and δ are referred to as the hyperparameters, and their estimates influence θ estimates. The hyperparameters are sampled from their corresponding prior distributions referred to as the hyperpriors, $\pi(\alpha)$, $\pi(\gamma)$, and $\pi(\delta)$, in the lowest stage of the hierarchy (see figure 6.4). The priors of the unknown parameters θ and β are estimated from these hyperparameters in the intermediate stage as shown in figure 6.4. The parameter estimates are then conditioned on the data to form the likelihood function in the highest stage of the hierarchy.

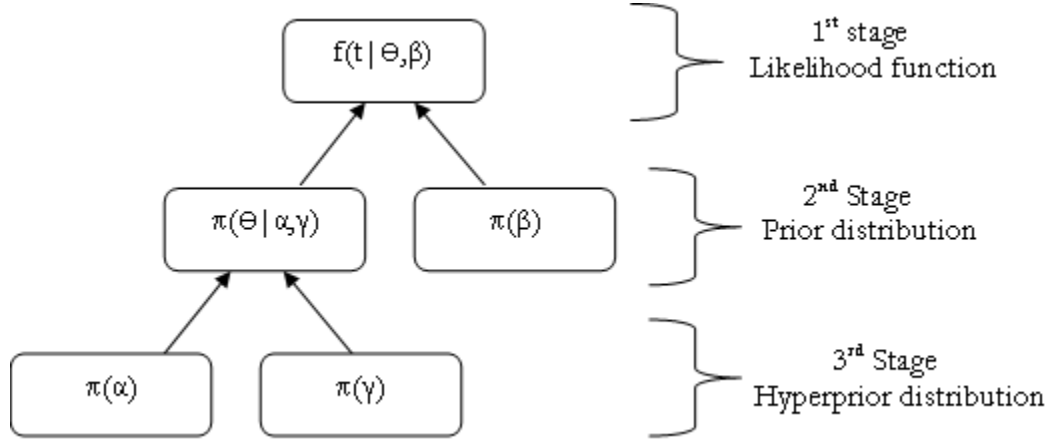


Figure 6.4. Graphical representation of the 3-stage hierarchical Bayesian framework

In equation 6.12, α depends on ΔH , k , and T . ΔH and k are constants for different dielectric materials. The test temperature, T should ideally be constant. However, variation usually occurs as a result of environmental conditions and calibration errors. In this analysis, we examine two temperature scenarios and the corresponding α values, namely: (i) constant temperature and α is constant at all stress levels and, (ii) varying temperature and α from one stress level to the other.

In equation 6.12, γ is the acceleration factor which determines the magnitude of the decrease in dielectric characteristic life θ , when the stress level is increased by one unit [34]. When γ is unknown, then it can be estimated by one of two methods: (i) the slope of the log-linear model in equation 6.12, or (ii) the ratio of dielectric characteristic life estimate at normal use conditions (using field data) and the characteristic life at stress condition (using data from accelerated failure tests) [34]. We will estimate the acceleration factor γ based on two conditions: when γ is constant at all stress levels, and when γ varies from one stress level to the other. The estimation error is assumed to be multiplicative.

Based on the foregoing discussion, we now construct two characteristic life models from equation 6.12 as follows [6]:

- Model I: When the dielectric characteristic life, θ_l , is different at every stress level, and the failure parameter α and the acceleration factor γ are constant at all stress levels as follows:

$$\ln(\theta_l) = \alpha - \gamma E_l + \delta_l \quad (6.13)$$

- Model II: When the characteristics life, θ_l , is different at every stress level, and both the failure parameter α_l and the acceleration factor γ_l differ at each stress level as follows:

$$\ln(\theta_l) = \alpha_l - \gamma_l E_l + \delta_l \quad (6.14)$$

where $(\cdot)_l$ denotes the l^{th} stress level.

The following notations are used to describe the implementation of the hierarchical Bayesian model.

- E_l : the l^{th} stress level (electric field) in MV/cm, for $l = 1, \dots, m$
- t_{lk} : the k^{th} failure time at the l^{th} stress level for $k = 1, \dots, n$
- t_l : the l^{th} vector of n_k failure times, for the n_k units at the l^{th} stress level
- T : the complete (n by m) matrix of all failure times t_{lk} for $l = 1, \dots, m$ and $k = 1, \dots, n$
- β : the shape parameter of the Weibull lifetime distribution, assumed equal at all stress levels
- θ_l : the scale parameter of the Weibull lifetime distribution under stress level l

- α : model I acceleration regression intercept
- γ : model I acceleration regression slope
- α_l : model II acceleration regression intercept at the l^{th} stress level
- γ_l : model II acceleration regression slope at the l^{th} stress level
- δ_l : the acceleration model error term at the l^{th} stress level

The hierarchical Bayesian model is implemented in stages. At the lowest level the hyperpriors, $\pi(\alpha)$, $\pi(\gamma)$ and $\pi(\delta)$ are specified. In the intermediate level, the unknown parameters (θ and β) are estimated from the hyperparameters, and at the highest level, the likelihood of the dielectric failure time conditioned on the estimated parameters is determined.

The hierarchical Bayesian models I and II are expressed as follows:

Model I

$$\begin{aligned}
f(t_{ik}|\theta_l, \beta) &\sim Weibull(\theta_l, \beta) \\
\ln(\theta_l|\alpha, \gamma, \delta_l) &\sim \alpha - \gamma E_l + \delta_l \\
\beta|a, b &\sim \pi_\beta(a, b) \\
\alpha|c, d &\sim \pi_\alpha(c, d) \\
\gamma|e, f &\sim \pi_\gamma(e, f) \\
\delta_l|\sigma^2 &\sim Normal(0, \sigma^2)
\end{aligned} \tag{6.15}$$

Model II

$$\begin{aligned}
f(t_{lk}|\theta_l, \beta_l) &\sim Weibull(\theta_l, \beta_l) \\
\ln(\theta_l|\alpha_l, \gamma_l, \delta_l) &\sim \alpha_l - \gamma_l E_l + \delta_l \\
\beta_l|a, b &\sim \pi_{\beta_l}(a, b) \\
\alpha_l|c_l, d_l &\sim \pi_{\alpha_l}(c_l, d_l) \\
\gamma_l|e_l, f_l &\sim \pi_{\gamma_l}(e_l, f_l) \\
\delta_l|\sigma^2 &\sim Normal(0, \sigma^2)
\end{aligned} \tag{6.16}$$

6.5.1 Limits of the Parameters and Hyperparameters

Hierarchical Bayesian models enable us to use a priori subjective information to set limits to the bounded uniform priors as follows:

1. Shape parameter:

The shape parameter β , is non-negative, and it determines the nature of the failure rate. We opt for $\pi(\beta) \sim \text{uniform}(0.0, 4.0)$.

2. Estimation error:

The estimation error δ , which is assumed multiplicative is elicited a standard normal, thus: $\pi(\delta) \sim \text{normal}(0.0, \sigma^2)$.

3. Characteristic life:

The dielectric characteristic life α , is a dimensionless parameter and is expressed as: $\alpha = \frac{\Delta H}{kT}$. The activation energy ΔH , is described as the amount of energy needed for a chemical reaction to take place. We will not get into the details of the contribution of atomic electronegativity towards the strength of covalent bonds in the dielectric bonds and the resultant bond weakening due to the

applied electric field. Instead, we will assume that for single-bonds of interest, the activation energy values range from ~ 0.1 eV to ~ 0.9 eV (for standard very-large-scale-integration (VLSI) gate-oxide processing [52], [102]. T is the test temperature which in our work will be assumed to range between 125 to 250 Kelvin [27], [103], and k is the Boltzmann constant which relates the energy at the atomic level with temperature at the bulk level [27]. This constant is given as 8.6173415×10^{-5} eV/K for application in semiconductor physics calculations [103]. These specifications result in α values ranging between ~ 4 and ~ 20 [103]. For implementation purposes, we extend the range of α from 1 to 50, that is $\pi(\alpha) \sim \text{uniform}(1, 50)$.

4. Acceleration factor:

The non-negative acceleration factor, γ will vary from 0.1 to 10, that is $\pi(\gamma) \sim \text{uniform}(0.1, 10)$ [52].

In summary, our model uses the following hyperprior distributions.

$$\pi(\beta) \sim \text{Uniform}(0.0, 4.0)$$

$$\pi(\alpha) \sim \text{Uniform}(1, 50)$$

$$\pi(\gamma) \sim \text{Uniform}(0.1, 10)$$

$$\pi(\delta) \sim \text{Normal}(0.0, 0.00001)$$

6.6 Posterior Computation

Given a vector of k unknown parameters $\theta_1, \dots, \theta_k$, the full or complete conditional distribution for each parameter is described by:

$$P(\theta_i | \theta_{j \neq i}, T) \text{ for } i, j \in m$$

For instance in our case, model I has the following unknown parameters: β , α , γ , and δ_l . In our implementation, we use an iterative mechanism that samples from the full conditional probability functions of each parameter, given the updated values of all the other parameters as follows [6]:

$$f(\beta|T, \alpha, \gamma, \Delta_l) \propto \pi(\beta) \prod_{l=1}^m \prod_{k=1}^{n_l} f(t_{lk}|\beta, \alpha, \gamma, \delta_l) \quad (6.17)$$

$$f(\alpha|T, \beta, \gamma, \Delta_l) \propto \pi(\alpha) \prod_{l=1}^m \prod_{k=1}^{n_l} f(t_{lk}|\beta, \alpha, \gamma, \delta_l) \quad (6.18)$$

$$f(\gamma|T, \beta, \alpha, \Delta_l) \propto \pi(\gamma) \prod_{l=1}^m \prod_{k=1}^{n_l} f(t_{lk}|\beta, \alpha, \gamma, \delta_l) \quad (6.19)$$

$$f(\delta_l|T, \beta, \gamma, \Delta_{-l}) \propto \pi(\delta_l) \prod_{k=1}^{n_l} f(t_{lk}|\beta, \alpha, \gamma, \delta_l) \quad (6.20)$$

where

$$\Delta_{-l} = \delta_1, \dots, \delta_{l-1}, \delta_{l+1}, \dots, \delta_m$$

and

$$f(t_{lk}|\beta, \alpha, \gamma, \delta_l) = \frac{\beta}{\theta} \left(\frac{t_{l,k}}{\theta} \right)^{\beta-1} \exp \left(-\frac{t_{l,k}}{\theta} \right) \quad (6.21)$$

where

$$\theta = \exp(\alpha + \gamma E_l + \delta_l)$$

6.6.1 MCMC Simulation

The convolution between the prior and the likelihood functions in order to obtain the posterior function cannot be determined in closed form due to its complexity. Therefore we used the MCMC approach as a means to obtain the posterior estimates. The simulations were implemented using the OpenBUGS programming language. OpenBUGS is a freely available statistical software, and it is an advancement to the

precursor software-WinBUGS, which stands for Windows Bayesian inference Using Gibbs Sampling [97]. The BUGS program implements Gibbs sampling approach, which is one of the famous Markov Chain Monte Carlo (MCMC) methods for solving intractable Bayesian problems. MCMC approaches are used to solve multiple integration required to obtain the marginal densities needed for Bayesian calculations. The objective of the MCMC methods is to create an ergodic Markov chain whose limiting or stationary distribution closely approximates the posterior distribution [104]. Generally, MCMC methods are implemented using two fundamental sampling algorithms namely the Metropolis-Hastings (M-H) and Gibbs sampling techniques [105]. More detailed discussions about MCMC sampling techniques as well as a rich reference of related texts are provided in [97], [104] and [106].

In OpenBUGS, an unknown random variable is called a node and its full conditional distribution is the probability density function conditioned on the updated values of all other stochastic nodes, that is, all other unknown parameters in the model [6]. For a given node, the full conditional density is proportional to the product of its prior density, and the likelihood function conditioned on the updated status of the all other nodes as shown in equations 6.18 to 6.20 [107]. WinBUGS does not necessarily require explicit evaluation of multiple integrals of the full conditional density functions. Instead, the computation is solved by continuously sampling from each posterior, until after a large number of iterations, the resulting k-tuple converges in distribution to a draw from the posterior distribution [97].

When sampling, the Gibbs sampler in the OpenBUGS software first attempts to recognize if the parameter samples are to be drawn from a conjugate prior. If conjugacy is fulfilled, the samples are drawn via direct sampling methodology using the prior distribution. If the prior is not conjugate, WinBUGS does the sampling numerically through the rejection sampling technique [97]. The most common rejection

sampling techniques are the Metropolis algorithm, adaptive rejection sampling (ARS) algorithm and slice sampling. The choice depends on the parameter descriptions. For instance, ARS is used when the parameter distribution is log-concave, slice sampling is employed when the support of the prior is bounded, while Metropolis algorithm is used when the support is unbounded [97], [108]. In our implementation, we let OpenBUGS run with the default sampling procedure appropriate for the prior distributions, which were: (i) interval slice sampling for α , β and γ and (ii) adaptive Metropolis-Hastings sampler for δ .

OpenBUGS also allows for simultaneous execution of multiple chains. A chain is described as a set of sampling initial values that the modeler defines, corresponding to each unknown parameter or node. While the ergodic nature of the Markov chain in the MCMC simulation ensures that the posterior solution is invariant to the initial points, the initial values affect convergence speed [109]. Such parallel chains are useful for model convergence diagnosis [110]. We implemented all simulations with two chains as a way to monitor the convergence.

The credibility of MCMC posterior results depends on model convergence. Basically, model convergence ensures that the iterative simulations reach an equilibrium state of the Markov Chain. Therefore, when the Markov chain induced by the MCMC algorithm fails to converge, the resulting posterior estimates will either be biased or unreliable [108]. OpenBUGS provides convergence diagnostics using trace plot, which are graphs of the updated posterior estimates at each iteration. In this work, convergence was determined by monitoring the trace plots, and this enabled us to decide on the simulation burn-in period. Figure 6.5 contains sample trace plots of β , γ and δ , and each shows the plots of parameter values at each iteration, for two separate chains (represented by the red and blue lines) at steady state. Based on the trial

simulations, we implemented 100,000 iterations with 20,000 burn-in iterations for model I and II.

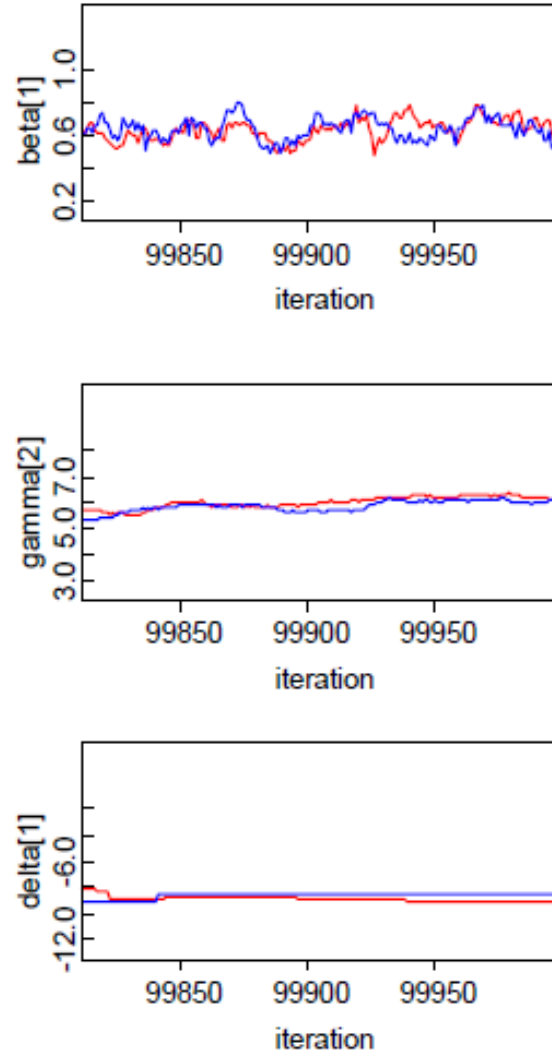


Figure 6.5. Trace plots indicating convergence

Figure 6.6 and 6.7 are schematic presentation of the hierarchical nature of models I and II respectively.

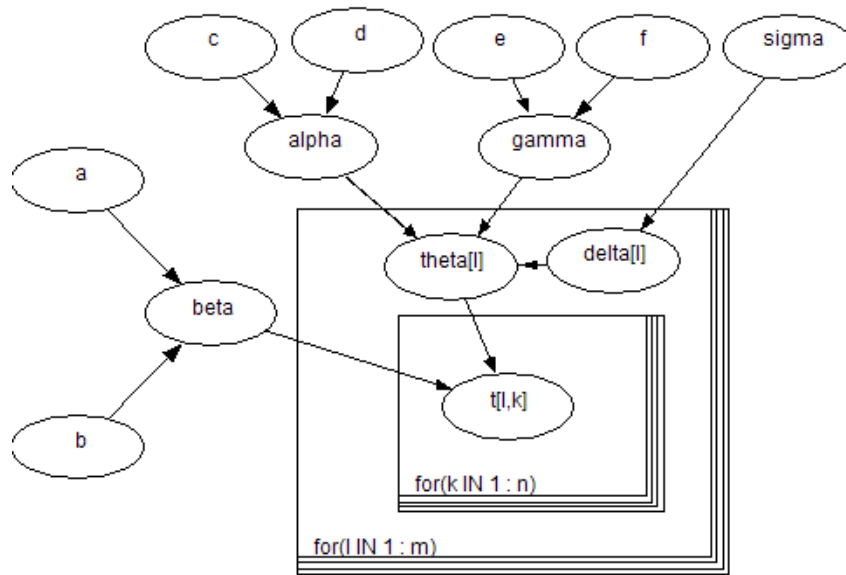


Figure 6.6. Schematic presentation of model I hierarchical Bayesian model

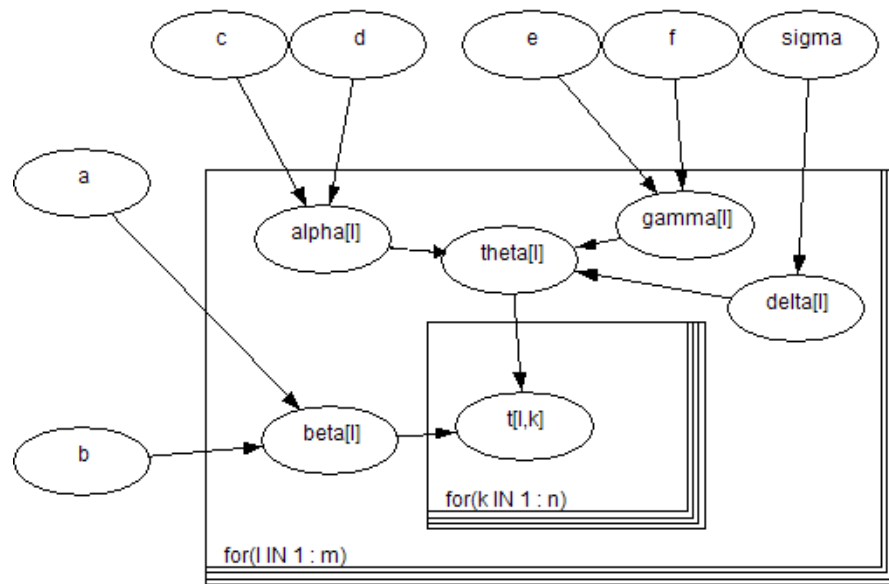


Figure 6.7. Schematic presentation of model II hierarchical Bayesian model

6.7 MCMC Simulation Results

When the samples are generated by an MCMC algorithm, the factors that influence the quality of point estimators include sampling error, estimation bias and systematic bias [108]. Sampling error, referred to as Monte Carlo (MC) error in MCMC arises from the difference between the posterior samples and the actual underlying distribution. In this work we used the MC errors to compare model I results against model II. The OpenBUGS results also give the 5 and 95 percentile limits of each parameter estimate. The best estimates should result in posterior estimates with the least standard deviation and least MC error.

The MCMC simulation results for models I and II are summarized in tables 6.4 and 6.5 respectively. In model I, we assume that the Weibull shape parameter, β and the Arrhenius-Weibull model parameters, α and γ are constant across all stress levels. In model II, we assume that β , α and γ different at each stress level. The following prior distributions were used for both models:

$$\begin{aligned}\pi(\beta) &\sim \text{Uniform}(0.0, 4.0) \\ \pi(\alpha) &\sim \text{Uniform}(1, 50) \\ \pi(\gamma) &\sim \text{Uniform}(0.1, 10) \\ \pi(\delta) &\sim \text{Normal}(0.0, 0.00001)\end{aligned}$$

6.7.1 Comparison of Hierarchical Bayesian Model I and II

The Weibull shape parameter β from Model I is 0.641 as seen in table 6.4. Model II gave several posterior mean values for β which are quite similar at a glance as seen in table 6.5. The β value at the 4th stress level is quite removed from the rest, possibly

Table 6.4. Posterior results:Bayesian hierarchical model I

Node	mean	st dev	MC error	5%	median	95%
alpha	24.66	11.67	0.489	2.485	25.520	45.230
beta	0.641	0.037	0.001	0.615	0.684	0.760
delta[1]	15.420	8.947	0.481	4.289	18.740	26.100
delta[2]	13.730	8.494	0.457	2.960	16.680	23.810
delta[3]	12.270	8.038	0.432	1.828	15.130	21.820
delta[4]	10.690	7.579	0.408	0.617	13.260	19.690
delta[5]	9.572	7.132	0.383	-0.220	11.830	18.100
delta[6]	8.245	6.681	0.359	-1.233	10.370	16.230
gamma	2.452	2.399	0.129	1.595	2.571	7.187
log theta[1]	5.903	0.266	0.003	5.394	5.898	6.438
log theta[2]	6.737	0.278	0.004	6.204	6.731	7.297
log theta[3]	7.202	0.257	0.004	6.71	7.197	7.721
log theta[4]	7.837	0.262	0.004	7.346	7.83	8.374
log theta[5]	7.837	0.260	0.004	7.332	7.835	8.358
log theta[6]	8.134	0.260	0.004	7.643	8.128	8.662
theta[1]	379.4	104.9	1.492	220.2	364.5	625.2
theta[2]	877.0	254.5	3.828	494.6	837.8	1476.0
theta[3]	1388.0	368.9	5.981	820.5	1335.0	2256.0
theta[4]	2624.0	720.5	11.22	1550.0	2515.0	4334.0
theta[5]	2621.0	704.7	10.89	1529.0	2527.0	4262.0
theta[6]	3529.0	956.8	15.23	2086.0	3388.0	5780.0

due to an outlier in the failure data at the 4th stress level. The mean posterior of β from model II, which is the average value of the estimates at 5 stress levels (excluding $\beta[4]$ is 0.640.

As indicated previously in section 6.3.3, β , the Weibull slope, is an indicator of a change in the failure mechanisms, and hence failure rate at varied stress levels. That is, a constant shape parameter across the stress levels indicates that there is no change in the failure mode as the stress level is increased and vice versa. Based on the estimates of β from the posterior results of model I and II, we conclude that there was no significant change in the failure mechanism of the dielectric at higher stress

Table 6.5. Posterior results:Bayesian hierarchical model II

Node	mean	st dev	MC error	5%	median	95%
alpha[1]	24.76	14.13	0.589	1.654	18.88	47.87
alpha[2]	22.56	12.03	0.500	1.923	20.69	43.2
alpha[3]	23.6	13.98	0.583	1.995	22.02	48.61
alpha[4]	25.59	12.56	0.523	4.969	30.18	49.34
alpha[5]	22.45	11.9	0.495	6.354	32.21	49.29
alpha[6]	25.08	12.81	0.534	2.539	27.48	48.38
beta[1]	0.641	0.086	9.85E-04	0.464	0.651	0.803
beta[2]	0.620	0.087	0.001	0.434	0.590	0.779
beta[3]	0.656	0.090	0.001	0.507	0.673	0.865
beta[4]	0.882†	0.119	0.002	0.660	0.878	1.13
beta[5]	0.643	0.088	0.002	0.501	0.659	0.844
beta[6]	0.632	0.090	0.002	0.524	0.69	0.878
delta[1]	3.918	8.139	0.341	-9.342	6.35	15.91
delta[2]	-0.656	5.321	0.223	-7.047	-2.35	11.21
delta[3]	9.569	16.73	0.702	-11.34	18.58	34.93
delta[4]	4.862	5.465	0.229	-3.346	3.136	15.11
delta[5]	-7.459	5.284	0.221	-16.79	-7.642	0.892
delta[6]	7.239	6.617	0.277	-1.456	5.394	24.47
gamma[1]	2.499	2.149	0.089	0.353	2.057	6.085
gamma[2]	2.121	1.56	0.064	0.566	2.972	6.039
gamma[3]	2.541	2.52	0.105	0.599	2.096	7.397
gamma[4]	2.265	1.684	0.070	1.798	2.808	7.583
gamma[5]	2.399	1.635	0.068	0.700	2.024	7.816
gamma[6]	2.429	1.741	0.072	1.992	2.667	8.728
log theta[1]	5.903	0.271	0.001	5.387	5.898	6.448
log theta[2]	6.734	0.278	0.001	6.205	6.728	7.299
log theta[3]	7.196	0.259	0.001	6.703	7.190	7.721
log theta[4]	7.845	0.259	0.001	7.353	7.839	8.372
log theta[5]	7.841	0.261	0.001	7.345	7.837	8.369
log theta[6]	8.128	0.256	0.001	7.639	8.122	8.648
theta[1]	379.9	106.9	0.481	218.6	364.3	631.6
theta[2]	874.2	254.1	1.056	495.2	835.1	1479.0
theta[3]	1380.0	370.6	1.568	815.1	1326.0	2255.0
theta[4]	2641.0	710.3	2.887	1561.0	2536.0	4323.0
theta[5]	2633.0	713.8	2.993	1548.0	2533.0	4313.0
theta[6]	3504.0	930.5	3.931	2077.0	3369.0	5696.0

levels. Also, results from model I and II show that the dielectric failure parameters α and γ are consistent across the stress levels. Following these results we assumed

that β , α and γ are constant across stress levels, and analyzed the sensitivity of the Bayesian hierarchical approach using model I.

6.8 Sensitivity Analysis

Possible sources of uncertainty in Bayesian models include the choice of the likelihood and prior functions, and the number of tiers in the hierarchical model [107]. Therefore the most appropriate method to check for model robustness is to carry out sensitivity analysis to determine whether the posterior estimates remain unchanged after slight perturbations in the prior and likelihood functions [97]. If the parameters remain the same, then the posterior is considered robust to a variety of data from the same problem domain.

In this work, posterior sensitivity was implemented by considering the different combinations of the prior distribution as shown table 6.6 for model I only. In this table, the base prior is the initial combination of prior distributions that was used to produce the results in tables 6.4 and 6.5. The results of the base prior are then compared to results from five other prior distribution combinations in table 6.7. The notation is such that $\alpha[1]$ denotes the posterior estimate of α using prior combination 1, and $\log\theta[11]$ denotes the posterior estimate of $\ln(\theta)$ at *stress level* [1] using *prior combination* [1].

Table 6.6. Table of priors for sensitivity analysis

Priors	$\pi(\alpha)$	$\pi(\gamma)$	$\pi(\beta)$	$\pi(\delta)$
Base (1)	uniform (1,50)	uniform(0.1,10)	uniform(0,4)	normal(0,0.000001)
2	uniform (1,100)	uniform(0.1,10)	uniform(0,4)	normal(0,0.000001)
3	normal (50,1/100)	normal(5,1/10)	uniform(0,4)	normal(0,0.000001)
4	normal (50,1/100)	uniform(0.1,10)	uniform(0,4)	normal(0,0.000001)
5	uniform (1,50)	normal(5,1/10)	uniform(0,4)	normal(0,0.000001)
6	uniform (10,40)	uniform(1,5)	uniform(0,4)	normal(0,0.000001)

The idea behind this sensitivity analysis is to determine the robustness of our hierarchical Bayesian model, by demonstrating the consistency of the posterior estimates of the Weibull shape parameter β , the dielectric characteristic life θ , and the failure parameters α and γ . Based on the results in table 6.7, all the posterior estimates maintain consistent values at all prior combinations. α and γ estimates for prior combinations 1 and 2 exhibit relatively higher standard deviations than those obtained with prior combinations 3,4, 5 and 6. This indicates that one should be careful in defining the limits of assigned uniform priors. Otherwise the low MC errors for β and $\ln(\theta)$ estimates indicate that the Bayesian model is robust to the assigned prior distributions.

Figures 6.8, 6.9, 6.10 and 6.11 show the posterior kernel densities for $\ln(\theta)$, β , α and γ respectively. Both $\ln(\theta)$ and β posterior kernels are similar at all prior combinations (1 to 6) as seen in figures 6.8 and 6.9. This indicates that the posterior estimates for the shape and characteristic life parameters are invariant to changes in the prior distributions. Though posterior estimates of α and γ remained constant (see table 6.7), figures 6.10 and 6.11 show some variability, which we attribute to the different boundaries that were set for their prior distributions.

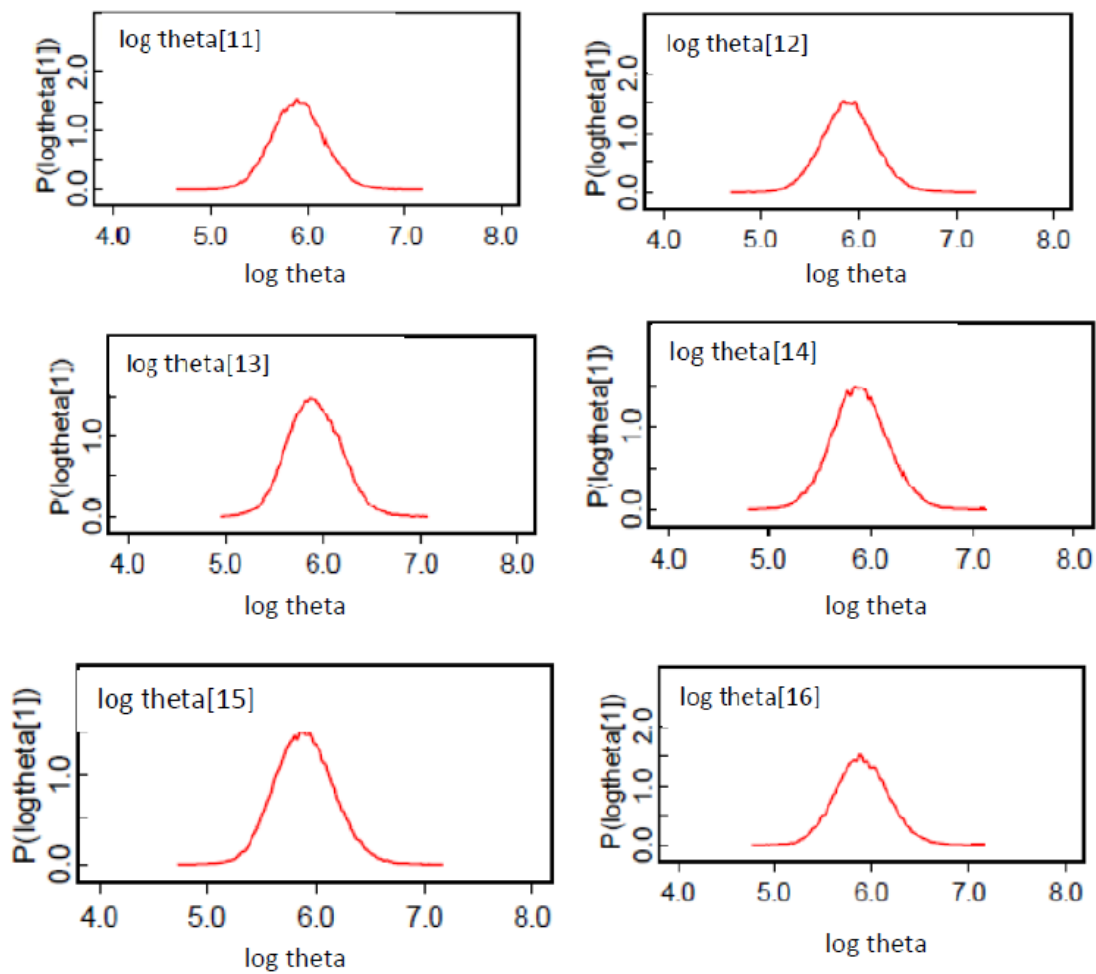


Figure 6.8. $\ln(\theta)$ posterior kernel density

Table 6.7. Sensitivity analysis

Node	mean	st dev	MC error	5%	median	95%	Prior
alpha[1]	24.66	11.67	4.90E-01	2.49	25.52	45.23	uniform (1,50)
alpha[2]	24.53	11.12	5.97E-01	10.45	27.31	44.88	normal (50,1/100)
alpha[3]	23.97	0.27	1.27E-02	22.44	23.99	24.45	normal (50,1/100)
alpha[4]	22.79	1.22	6.09E-02	19.78	22.78	24.57	normal (50,1/100)
alpha[5]	24.53	4.10	2.05E-01	11.87	14.76	26.09	uniform (1,50)
alpha[6]	22.54	8.56	4.06E-01	10.23	20.04	36.70	uniform (10,40)
beta[1]	0.68	0.04	6.15E-04	0.61	0.68	0.76	uniform(0,4)
beta[2]	0.69	0.04	1.01E-03	0.61	0.68	0.76	uniform(0,4)
beta[3]	0.68	0.04	1.55E-03	0.61	0.68	0.76	uniform(0,4)
beta[4]	0.68	0.04	9.47E-04	0.61	0.68	0.76	uniform(0,4)
beta[5]	0.69	0.04	8.64E-04	0.61	0.68	0.76	uniform(0,4)
beta[6]	0.69	0.04	7.91E-04	0.61	0.68	0.76	uniform(0,4)
gamma[1]	2.32	1.60	6.71E-02	1.27	3.92	7.09	uniform(0.1,10)
gamma[2]	2.46	2.40	1.29E-01	2.60	6.57	9.19	uniform(0.1,10)
gamma[3]	2.24	0.25	1.24E-02	1.37	3.00	3.20	normal(5,1/10)
gamma[4]	2.93	0.85	4.27E-02	2.48	4.98	6.00	uniform(0.1,10)
gamma[5]	2.37	0.28	9.11E-03	3.05	3.33	3.73	normal(5,1/10)
gamma[6]	2.60	0.95	4.49E-02	2.11	3.60	4.96	uniform(1,5)
log theta[11]	5.90	0.27	3.87E-03	5.39	5.90	6.438	
log theta[21]	5.90	0.27	3.98E-03	5.38	5.90	6.439	
log theta[31]	5.91	0.27	5.01E-03	5.40	5.91	6.448	
log theta[41]	5.90	0.27	5.42E-03	5.37	5.89	6.44	
log theta[51]	5.90	0.27	4.11E-03	5.39	5.89	6.44	
log theta[61]	5.91	0.27	4.72E-03	5.39	5.90	6.45	
log theta[12]	6.74	0.28	4.18E-03	6.20	6.73	7.30	
log theta[22]	6.74	0.28	4.32E-03	6.21	6.73	7.31	
log theta[32]	6.75	0.27	5.00E-03	6.23	6.74	7.30	
log theta[42]	6.74	0.28	5.17E-03	6.20	6.73	7.30	
log theta[52]	6.74	0.28	4.46E-03	6.21	6.73	7.29	
log theta[62]	6.73	0.28	4.96E-03	6.20	6.72	7.29	
log theta[13]	7.20	0.26	4.13E-03	6.71	7.20	7.72	
log theta[23]	7.19	0.26	4.28E-03	6.69	7.19	7.72	
log theta[33]	7.20	0.26	4.94E-03	6.71	7.20	7.74	
log theta[43]	7.20	0.25	4.87E-03	6.73	7.19	7.71	
log theta[53]	7.20	0.26	3.99E-03	6.70	7.19	7.721	
log theta[63]	7.20	0.26	4.71E-03	6.71	7.19	7.71	
log theta[14]	7.84	0.26	4.12E-03	7.35	7.83	8.37	
log theta[24]	7.85	0.26	4.24E-03	7.36	7.84	8.39	
log theta[34]	7.85	0.26	4.49E-03	7.36	7.84	8.37	
log theta[44]	7.85	0.26	5.28E-03	7.37	7.84	8.38	
log theta[54]	7.85	0.26	3.91E-03	7.36	7.84	8.37	
log theta[64]	7.84	0.26	4.86E-03	7.35	7.83	8.374	
log theta[15]	7.84	0.26	4.10E-03	7.33	7.84	8.36	
log theta[25]	7.84	0.26	4.11E-03	7.34	7.83	8.36	
log theta[35]	7.85	0.26	4.58E-03	7.36	7.85	8.38	
log theta[45]	7.84	0.26	5.03E-03	7.33	7.84	8.35	
log theta[55]	7.84	0.26	4.00E-03	7.33	7.83	8.36	
log theta[65]	7.84	0.26	4.80E-03	7.35	7.84	8.37	
log theta[16]	8.13	0.26	4.18E-03	7.64	8.13	8.662	
log theta[26]	8.13	0.26	4.27E-03	7.64	8.12	8.64	
log theta[36]	8.13	0.25	4.36E-03	7.65	8.13	8.64	
log theta[46]	8.13	0.26	5.27E-03	7.63	8.12	8.65	
log theta[56]	8.12	0.26	4.23E-03	7.64	8.11	8.65	
log theta[66]	8.13	0.26	4.65E-03	7.64	8.12	8.65	

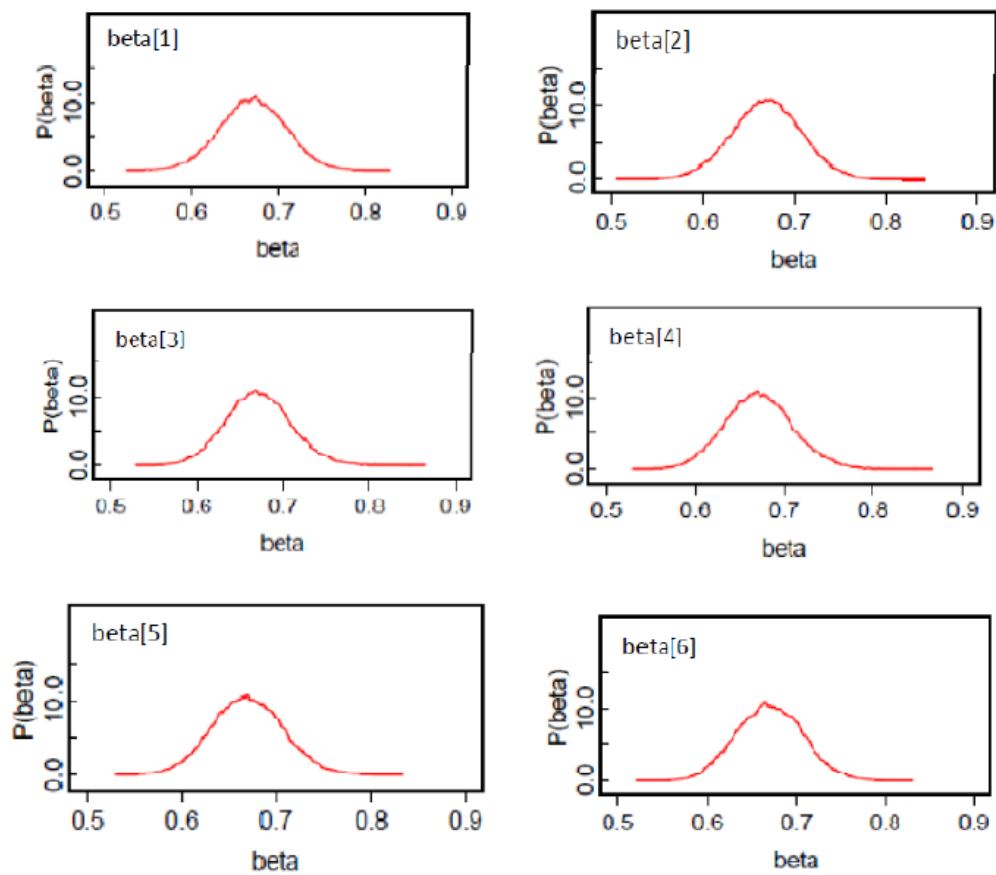


Figure 6.9. β posterior kernel density

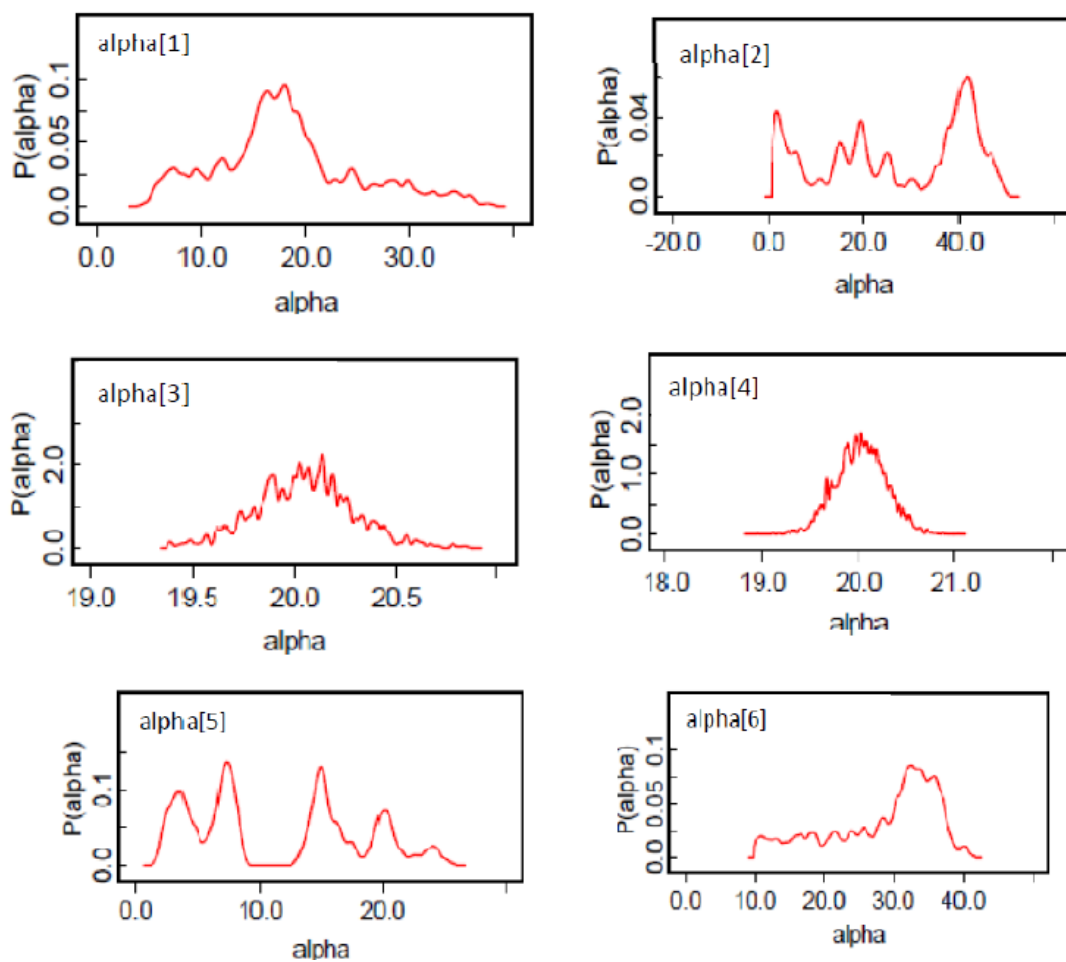


Figure 6.10. α posterior kernel density

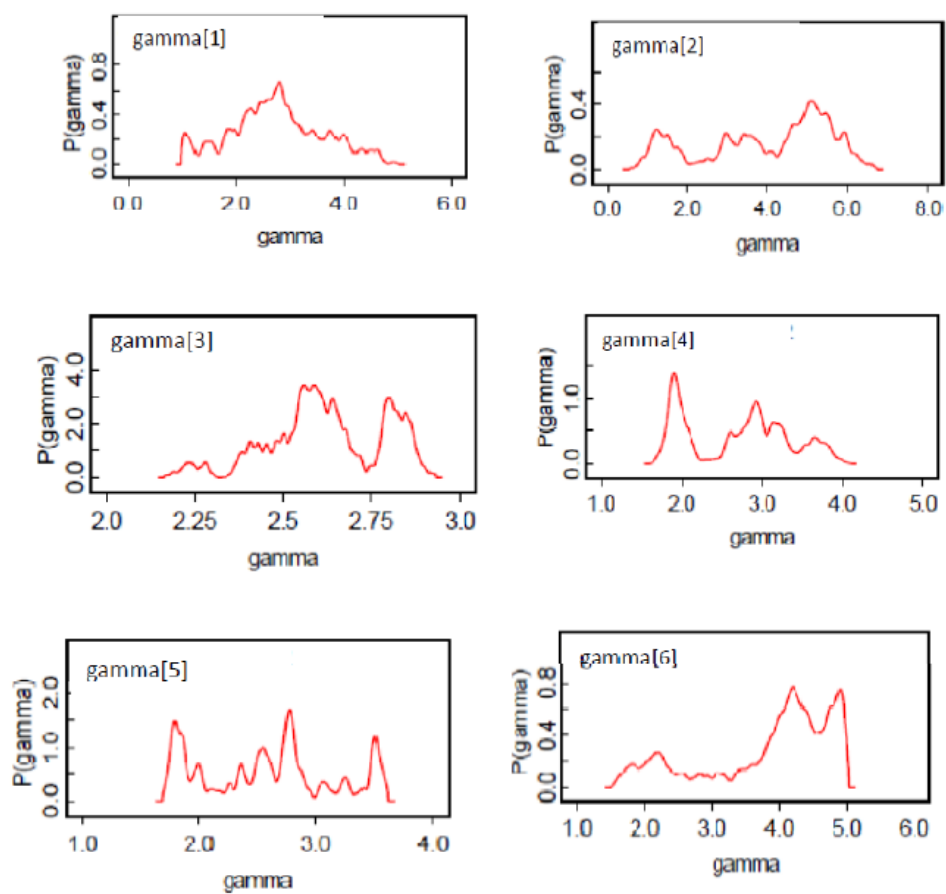


Figure 6.11. γ posterior kernel density

6.9 Reliability Inference

The third objective of this dissertation as enumerated in chapter 3, is to make reliability inference at the accelerated test levels followed by reliability projections at operating conditions. To do so, the estimates of the dielectric characteristics life, and the Weibull shape parameter from the hierarchical Bayesian approach using model I and prior combination 1 of table 6.6 are compared to the corresponding estimates obtained using classical maximum likelihood procedure and least square estimation procedure based on median-rank regression.

6.9.1 Weibull Maximum Likelihood Estimates (MLE)

Given the failure data $T = t_{lk}$ for $l = 1, 2, \dots, m$ and $k = 1, 2, \dots, n_l$, the Maximum Likelihood Estimates of the shape parameters, β_l , and the dielectric characteristic life, θ_l at each stress level E_l were determined using MATLAB, resulting in a vector of θ_{MLE} and β_{MLE} at each stress level.

6.9.2 Weibull Least Square Error (LSE) Estimates

The Weibull LSE estimates were determined using the median-rank approach. In this method, the Weibull empirical cumulative distribution function (ECDF) is used to estimate the Weibull cumulative distribution function (CDF). The Weibull cumulative distribution function (CDF) is expressed as:

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\theta_l}\right)^{\beta_l}\right) \quad (6.22)$$

When the CDF, $F(t_{lk})$ is replaced by the ECDF, $\hat{F}(t_{lk})$, and double natural logarithm taken on both side of equation 6.22, we get:

$$\ln \left(\ln \left(\frac{1}{1 - \hat{F}(t_{lk})} \right) \right) = \beta_l \ln(t_{lk}) - \beta_l \ln(\theta_l) \quad (6.23)$$

Based on equation 6.23, the Weibull LSE estimation procedure is as follows:

1. Rank the failure times data in increasing order, $t_{l(1)}, \dots, t_{l(n_l)}$ for $l = 1, \dots, m$.
2. Determine the Weibull ECDF using the median rank estimation procedure as follows:

$$F(t_{lk}) = \frac{k - 0.3}{n + 0.4} \quad k = 1, \dots, n \quad l = 1, \dots, m \quad (6.24)$$

where n is the sample size and k is the rank.

3. Compute the response variable y such that:

$$y = \ln \left(\ln \left(\frac{1}{1 - F(t_{lk})} \right) \right) \quad (6.25)$$

4. Regress y in equation 6.25 against $\ln(t_{lk})$ using the following model:

$$y = a_l(\ln t_{lk}) + b_l \quad (6.26)$$

Relating equations 6.23 and 6.26, the Weibull shape parameter in equation 6.2 is equivalent to the regression slope of equation 6.23. The dielectric characteristic life is determined from the regression intercept and slope of equation 6.26 as follows:

$$\beta_l \ln(\theta_l) = b_l \implies \ln(\theta_l) = -\frac{b_l}{\beta_l} \quad (6.27)$$

The results of LSE is also a vector of size m , whose elements are θ_{LSE} and β_{LSE} at each stress level.

Tables 6.8 and 6.9 show a summary of the characteristic life θ , and shape parameter β at each of the stress levels E_l , for each estimation procedure (Bayes, MLE and LSE).

Table 6.8. Dielectric characteristics life estimates (seconds)

Stress	Methodology	Mean Estimate	5% limit	95% limit
8.1	Hierarchical Bayes	3.07E+02	2.94E+02	8.44E+02
	MLE	4.66E+02	1.56E+02	5.59E+02
	LSE	5.05E+02	4.43E+02	5.51E+02
7.9	Hierarchical Bayes	4.80E+02	4.54E+02	1.29E+03
	MLE	7.32E+02	2.52E+02	8.75E+02
	LSE	7.92E+02	6.96E+02	8.62E+02
7.7	Hierarchical Bayes	7.49E+02	7.02E+02	1.98E+03
	MLE	1.15E+03	4.05E+02	1.37E+03
	LSE	1.24E+03	1.09E+03	1.35E+03
7.5	Hierarchical Bayes	1.17E+03	1.09E+03	3.04E+03
	MLE	1.81E+03	6.52E+02	2.14E+03
	LSE	1.95E+03	1.72E+03	2.11E+03
7.3	Hierarchical Bayes	1.83E+03	1.68E+03	4.66E+03
	MLE	2.84E+03	1.05E+03	3.35E+03
	LSE	3.06E+03	2.70E+03	3.30E+03
7.1	Hierarchical Bayes	2.86E+03	2.60E+03	7.15E+03
	MLE	4.46E+03	1.69E+03	5.25E+03
	LSE	4.79E+03	4.25E+03	5.17E+03

Table 6.9. Weibull shape parameter estimates

Stress	Methodology	Mean Estimate	5% limit	95% limit
8.1	Hierarchical Bayes	0.624	0.464	0.803
	MLE	0.599	0.553	0.645
	LSE	0.599	0.553	0.645
7.9	Hierarchical Bayes	0.594	0.435	0.779
	MLE	0.583	0.437	0.778
	LSE	0.537	0.504	0.570
7.7	Hierarchical Bayes	0.676	0.507	0.865
	MLE	0.667	0.511	0.871
	LSE	0.633	0.596	0.671
7.5	Hierarchical Bayes	0.882	0.661	1.130
	MLE	0.871	0.666	1.139
	LSE	0.790	0.744	0.836
7.3	Hierarchical Bayes	0.663	0.501	0.845
	MLE	0.655	0.503	0.853
	LSE	0.625	0.597	0.653
7.1	Hierarchical Bayes	0.693	0.524	0.878
	MLE	0.683	0.527	0.884
	LSE	0.686	0.645	0.727

6.9.3 Extrapolating Dielectric Characteristic Life from Accelerated Level to Normal Use Condition

Based on the Arrhenius-Weibull failure model, (see section 6.3.2.1) the linearized relationship between the dielectric characteristic life and the stress level is given by:

$$\ln(\theta_l) = c + d(E_l) \quad (6.28)$$

The following linearized extrapolation models were developed using the estimates in table 6.8:

$$\ln(\widehat{\theta_{Bayes}}) = 23.65 - 2.156(E_l) \quad (6.29)$$

$$\ln(\widehat{\theta_{MLE}}) = 24.45 - 2.261(E_l) \quad (6.30)$$

$$\ln(\widehat{\theta_{LSE}}) = 24.40 - 2.250(E_l) \quad (6.31)$$

6.9.4 Mean Time to Failure (MTTF) Extrapolation

One of the measures of reliability is the Mean Time To Failure (MTTF), defined as the expected time to failure for non-repairable system, or the expected time between two consecutive failures for repairable systems [48]. Generally, given a vector of failure times, $T = \{t_1, \dots, t_n\}$, the MTTF estimate is given by:

$$M\widehat{FFT} = \int_0^\infty tf(t) dt \quad (6.32)$$

For the Weibull distribution, the MTTF estimate is given by [48]:

$$M\widehat{FFT} = \hat{\theta} \frac{1}{\hat{\beta}} \Gamma\left(\frac{1}{\hat{\beta}}\right) = \hat{\theta} \Gamma\left(1 + \frac{1}{\hat{\beta}}\right) \quad (6.33)$$

where $\hat{\theta}$ and $\hat{\beta}$ are the Weibull shape and characteristic life parameter estimates.

In this section, a comparison is made between the MTTF values resulting from the Bayesian model, the MLE and LSE estimation methods. Since the main objective is to extrapolate the MTTF from test conditions to normal use conditions, we use equations 6.34 to 6.36 (derived from equations 6.29 to 6.31) to compute the characteristic life estimates θ_l at stress levels that represent normal use conditions, approximately ranging from 2 MV/cm to 6 MV/cm [27].

$$\widehat{\theta_{(Bayes)}} = \exp(23.97 - 2.239E_l) \quad (6.34)$$

$$\widehat{\theta_{(MLE)}} = \exp(24.45 - 2.26E_l) \quad (6.35)$$

$$\widehat{\theta_{(LSE)}} = \exp(24.45 - 2.25E_l) \quad (6.36)$$

The Weibull shape parameter β for Bayes, MLE and LSE are averaged over all the β estimates at 8.1, 7.9, 7.7, 7.3, and 7.1 MV/cm stress levels. The value of β at 7.5 MV/cm stress level is remarkably high, perhaps due to an outlier in the failure data, and hence it is excluded from the averaging procedure. The corresponding mean β estimates are $\widehat{\beta}_{Bayes} = 0.65$, $\widehat{\beta}_{MLE} = 0.64$, and $\widehat{\beta}_{LSE} = 0.62$. These β estimates, together with θ estimates are used to calculate the MTTF values at the accelerated conditions. The results are summarized in table 6.10.

Table 6.10. Comparison of MTTF estimates

Stress level (MV/cm)	MTTF (in seconds)		
	Bayes	MLE	LSE
7.1	3.91E+03	6.20E+03	6.91E+03
7.3	2.50E+03	3.95E+03	4.42E+03
7.5	1.60E+03	2.52E+03	2.82E+03
7.7	1.02E+03	1.60E+03	1.79E+03
7.9	6.56E+02	1.02E+03	1.14E+03
8.1	4.19E+02	6.48E+02	7.29E+02

Table 6.11 and 6.12 show the Bayes, MLE and LSE 95% confidence interval of the extrapolated MTTF at 6 and 2 MV/cm. These confidence limits are also shown graphically in figure 6.12.

Table 6.11. Extrapolated MTTF 95% confidence interval at 6 MV/cm

	MTTF estimate (hours)	5% limit	95% limit
Bayes	1.26E+01	1.65E+01	2.30E+01
MLE	2.07E+01	1.24E+01	1.92E+01
LSE	2.29E+01	2.23E+01	2.31E+01

Table 6.12. Extrapolated MTTF 95% confidence interval at 2 MV/cm

	MTTF estimate (hours)	5% limit	95% limit
Bayes	1.16E+05	9.41E+04	1.18E+05
MLE	1.69E+05	1.50E+05	1.75E+05
LSE	1.85E+05	1.80E+05	1.88E+05

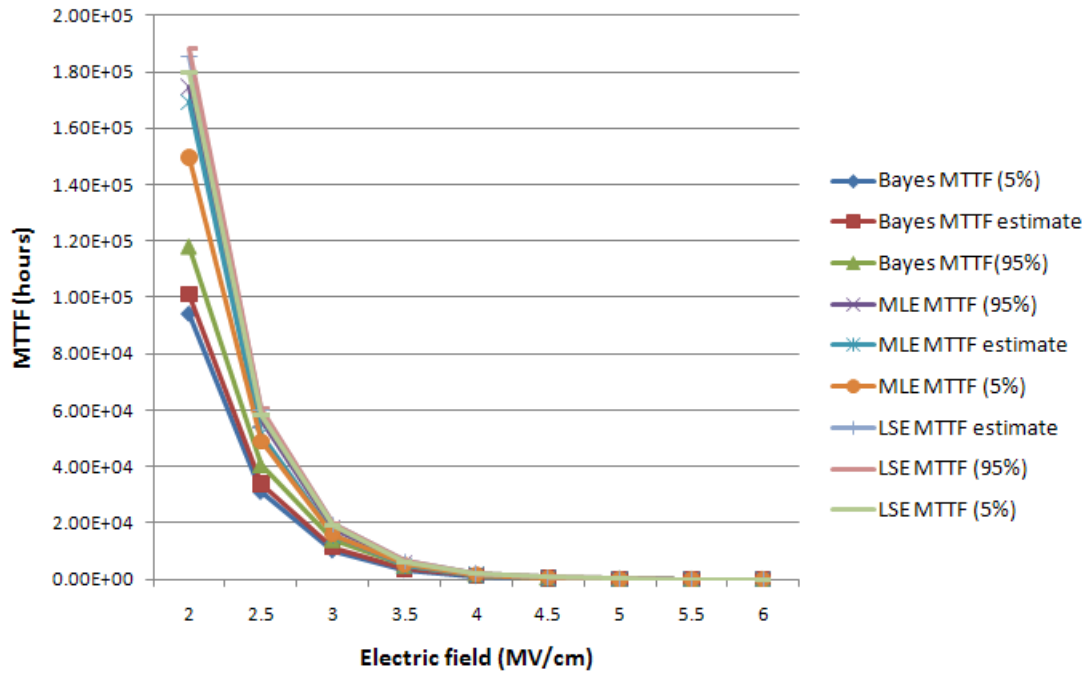


Figure 6.12. Graph of extrapolated MTTF

To make reliability inference from the extrapolated MTTF values, we assume that the Arrhenius-Weibull model is valid at low stress levels, representing normal use conditions. In figure 6.12, Bayes, MLE and LSE MTTF plots indicate that the mean time to failure decreases exponentially with increasing stress levels, and these results agree with the Arrhenius acceleration model. Though LSE and MLE give tighter 95% confidence interval than the Bayes approach, the MLE and LSE estimates are much higher, and thus optimistic than the Bayes approach. This difference is probably due to the small size of failure data that was used. MLE and LSE procedures are mostly advocated for because of their simplicity and asymptotic distribution optimality properties which is realized when there is plenty of data [111]. However, in situations when data is inadequate, which is typical of accelerated life tests, the Bayesian technique performs better.

CHAPTER 7

RESULTS AND MAJOR CONTRIBUTIONS

In this work, we develop a framework to study and analyze the behavior, properties, and failure mechanisms of nanoscale dielectric thin films. The framework includes a 3D simulation of the failure of dielectric thin films with the aim of understanding their failure characteristics. Both parametric and nonparametric techniques are then used to conduct reliability inference analyses of the dielectric film.

In this chapter specifically, we provide a summary of the results of our work. In addition, we discuss the major contributions of this research with a focus on the following three areas, namely: (i) use of simulation to characterize the dielectric failure process, (ii) failure density estimation using the kernel density approach, and (iii) reliability inference using Bayesian methodology.

7.1 3D Simulation Model

The 3D failure simulation model presented in chapter 4 includes a search algorithm that identifies the critical number of defects at breakdown, N_{bd} , as a function of the size of defects and the dielectric thickness. In the simulation, the cells are randomly occupied by defects, which is accomplished by switching of the occupied cells from a non defective to a defective state. Breakdown occurs when a collection of defects form conductive paths that bridge the dielectric layer. We use an empirical power function which relates the critical defect density at breakdown to the stress time as proposed by [62], [64], and [66], to determine realistic failure times.

The results in figure 4.14 show that defect densities ranging from 0.0001 to 0.0040 defects/ nm^3 were realized when the mean number of defects introduced into the model was fixed at 1 for the 1 nm and 5 nm simulated dielectric thickness respectively. A similar range, (0.0003 to 0.0046 defects/ nm^3) resulted when the average number of defects introduced was increased to 10. However, the range remarkably increased, (0.025 to 0.0534 defects/ nm^3) when the number of defects introduced at each simulation was further increased to 100. In addition, figures 4.15 to 4.17 show a consistent increase in dielectric breakdown time with increasing thickness.

Based on our results, we can draw the following conclusions:

1. The critical defect density and breakdown time increase with dielectric thickness, and the relationship between the critical defect density and dielectric thickness is possibly a power function.
2. The simulation is sensitive to the definition of cell communication, and it appears that among all choices, the path formation that uses 26-nearest neighbor cells most closely approximates the actual dielectric failure process.
3. The mean number of defects introduced at each iteration largely affects the critical defect density.
4. The variability of failure times is about the same at each dielectric thickness level. This is an indication that our simulation model is quite robust in predicting the relationship between the time to failure and dielectric thickness.

We compared our model results with those from the analytical model proposed by Sune et al. [67] that relates critical defect density to dielectric thickness. Our results (see figure 4.14) are in agreement with the analytical model, particularly with 26 degrees-of-communication (26-nearest neighbor cells), and with an average of 10

defects being introduced into the model at each iteration. Based on the Weibull plots (see figure 6.1) we can deduce that: (i) the dielectric characteristic life (the time at which the probability of failure for the device is 63.2%, assuming a Weibull underlying probability distribution) increases with dielectric thickness, and (ii) the rate of failure, (which is dependent on the Weibull slope or shape parameter) increases with time. With respect to the dielectric films, this means that it takes longer for failures to occur in thick films than in thin ones.

7.2 Kernel Density Estimation

In chapter 5, we used the kernel density approach to construct the probability density structure of actual dielectric failure data. To select a suitable variable bandwidth for the data, we optimized a cost function derived from the Mean Integrated Square Error (MISE). The results (see figures 5.2, 5.3, and Appendix A), show that the overall density structures using data-specific variable bandwidths and constant bandwidths were quite similar. However, variable bandwidths provide more detailed information of the underlying probability structure that govern the dielectric failure process. One drawback of the variable bandwidth is its sensitivity to the number of data points at each estimation interval. Consequently, it appears that a variable bandwidth would perform better with large data sets because it provides enough data points for effective smoothing at each interval within the data range.

7.3 Bayesian Inference

The Bayesian inference framework was constructed from the bounded uniform prior distribution, and the two-parameter Weibull likelihood function. Using the Arrhenius-Weibull relationship, we developed two models that relate the dielectric

characteristic life θ , the stress level (E_t), the acceleration factor γ , and the material and temperature related parameter α . In the first model (Model I, see chapter 6) α and γ are assumed constant at all stress levels, but in the second model (Model II) α and γ are assumed to differ at each stress levels. The optimal estimates of α and γ are used to update the estimates of the unknown parameters of the posterior distribution.

The results in tables 6.4 and 6.5 show that for stress levels (electric field) ranging from 8.1 to 7.1 MV/cm, the dielectric characteristic life, θ ranges from 380 to 3530 seconds respectively for both models. The Weibull shape parameter, β for model I is 0.641, and the average is 0.638 for model II. The material related failure parameter α , is 24.66 for model I and the average is 24.00 in model II. The acceleration factor, γ is 2.452 cm/MV for model I and the average is 2.375 cm/MV in model II.

Based on the simulation results we can conclude that:

1. The material and temperature related constant α , and the acceleration factor γ , remain constant across the stress levels, and this means that model I is a sufficient representation of the dielectric failure process.
2. The dielectric characteristic life, θ consistently increases with increasing stress levels, for both model I and II, meaning that failure occurs faster at accelerated stress levels. Specifically, the 95% confidence interval of characteristic life of a 2 nm thick high-k dielectric at accelerated stress levels, ~ 8.1 Mv/cm is 294 to 844 seconds, and the corresponding confidence interval at normal use condition, ~ 2 MV/cm is 1.75×10^8 to 3.85×10^8 seconds.
3. The Weibull shape parameter β , for a fixed dielectric thickness remains constant across the stress level and this physically means that an increase in the stress levels does not induce a different failure mechanism.

To perform reliability inference at normal use conditions using the accelerated data, we construct a log-linear extrapolation model to determine the dielectric characteristic life θ , and hence the Mean Time to Failure (MTTF) at normal use condition. Based on the MTTF analysis, the Bayesian approach results in the smallest mean time to failure in comparison to the MLE and LSE approaches. However, the MLE and LSE produce narrower 95% confidence intervals. Typically, the MLE and LSE estimation techniques have better asymptotic properties and therefore perform better when the available data is large. When there is insufficient data such as, in our case, the Bayesian estimation approach performs better. Moreover, by using the Bayesian method, we are able to incorporate subjective dielectric failure information to model the characteristic life parameter, which is not possible with the MLE and LSE techniques.

7.4 Contributions

The major contributions of this research are on three fronts, namely: (i) the use of simulation to characterize the dielectric failure process, (ii) failure density estimation using Kernel density approach, (iii) reliability inference using Bayesian methodology.

1. Failure Simulation:

Our contribution is in the development of an algorithm for a 3D simulation of dielectric failure which is available for users. Most of the dielectric failure simulations in the literature consider 3D models with dielectric films $\geq 10nm$ and assume that dielectric breakdown is triggered by the formation of the first conductive path that bridges the dielectric thickness [26], [57], [53]. In our research however, we consider much smaller thickness, that is, dielectric thickness ranging from 1 to 5 nm. We also conducted a sensitivity analysis of the critical

defect density and the failure times on the mean number of defects introduced at each simulation trial. In addition, we studied the effect of the number of critical conductive paths on the critical defect density as well as the failure times. The analysis of the failure times shows that the algorithm generates failure times that have consistent variability at all dielectric thickness levels.

2. Nonparametric Failure Density Estimation:

We develop a nonparametric methodology to estimate the dielectric failure density structure using the kernel approach with a variable bandwidth. While this technique has been used to successfully determine the underlying density structures for decades, our contribution in this regard is in the development of an optimal variable bandwidth based on a cost minimization algorithm using the Mean Integrated Square Error (MISE) for a given set of data. Previously, researchers have used the MISE to select a constant global bandwidth for an entire data range. Our findings show that when compared to a constant bandwidth, the variable bandwidth is sensitive to peaks within the data, and provides appropriate smoothing that ensures that details within the data structure are not masked.

3. Bayesian Reliability Inference:

Our contribution here is in the development of an integrated hierarchical Bayes model for dielectric reliability inference. Hierarchical Bayesian models have successfully been used for survival inference in public health and related research. We extend this application to dielectric failure analysis by proposing 3-stage model that incorporates the Arrhenius-Weibull relationship. Previous statistical models have used either classical parametric or nonparametric methods to model dielectric failure. In our case, we incorporate physics-of-failure models

to estimate failure parameters using the Hierarchical Bayesian methodology. The parameters, which include the dielectric characteristic life, the failure rate and the acceleration factor, are all necessary to predict the life of a dielectric thin film. The proposed framework can be adopted to determine the reliability inference of other nanomaterials and devices, possibly by replacing the failure distribution and physics-of-failure models with those related to the device of interest.

CHAPTER 8

SUGGESTIONS FOR FUTURE RESEARCH

The overarching goal of this work is to develop a framework for the reliability of nanoscale dielectric films. In pursuit of this goal, we have tried to provide modeling and analytical clarity regarding the inherent failure characteristics and hence the reliability of high-k gate dielectrics. The ultimate objective is to extend this framework to other nanoscale materials and devices so as to optimally characterize, predict and manage their reliability. Such an effort requires an interdisciplinary approach in order to understand both the physical relationships as well as the probabilistic and statistical complexities that underlie device and material behavior. While we have attempted to provide some clarity to some of these issues, several vexing questions still remain.

8.1 Dielectric Failure Simulation

We made several assumptions in the construction of the 3D simulation of dielectric failure. First, we assumed that the defects are spherical and that each defect fills up one cubic lattice in the simulation. An important question is to what extent does the change in the defect shape and dimension affect the critical defect density values? Also, what is the impact of overlapping defects (defects that span more than one cell) on the results of the simulation experiment and specifically the defect path formation? There is still controversy regarding not just the shape of the defect, but also the dimension and orientation of defects within the dielectric bulk and interfacial

region. A better understanding of these issues would make for better simulation and hence more realistic results of the dielectric failure.

Secondly, we assume that a breakdown occurs when the cluster of defects form five or more conductive paths that bridge the dielectric thickness. Previously, experts have modeled dielectric failure to occur when the first path is formed. Based on our findings the critical defect density when five or more paths are formed closely approximates actual dielectric failure behavior. Recently, experts have noted that not all defects equally contribute to path formation, and not all paths are available for electron transfer [63]. Therefore, the question of path efficiency requires special attention, more so, to draw a distinction between hard and soft dielectric failures.

8.2 Nonparametric Density Estimation

Nonparametric density estimation methods are useful for reliability inference, especially for failure phenomena whose data do not conform to traditional probability distributions. In this study, we use the Gaussian kernel to develop a Mean Integrated Squared Error (MISE) minimization algorithm for selecting a vector of optimal variable bandwidths for a given data. Our findings show that variable bandwidths from the AMISE-based cost minimization approach perform better with large data. Given that Bayesian approaches are useful in the case of small data samples, attempts should be made to determine Bayesian optimized variable bandwidths.

8.3 Bayesian Reliability Inference

In this study, we develop a 3-stage hierarchical Bayesian model to estimates dielectric failure parameters, using the two-parameter Weibull likelihood function. We also use a noninformative prior, given the uncertainty regarding the unknown parameter.

Though Bayesian approaches are useful when the data sample is small, the posterior structure is largely affected by the prior distribution. Therefore, more attention should be given to the choice of the prior that would possibly lead to a posterior distribution whose structure can be evaluated in closed form.

REFERENCES

- [1] M. Houssa. *Higk-K Gate Dielectrics*. Institute of Physics, Philadelphia, 2004.
- [2] J. Hicks, D. Bergstrom, M. Hattendorf, J. Jopling, J. Maiz, S. Pae, and C. Prasad. 45nm transistor reliability. *Intel Technology Journal*, 12:1–16, 2008.
- [3] J. Sune, E.Y. Wu, D. Jimenez, and Lai W. L. Statistics of soft and hard breakdown in thin SiO₂ gate oxides. *Microelectronic Reliability*, 43:1185–1192, 2003.
- [4] S.L. Jeng, J.C. Lu, and K. Wang. A review of reliability research on nanotechnology. *IEEE Transactions on Reliability*, 56:401–410, 2007.
- [5] E. A. Elsayed. *Reliability Engineering*. Addison Wesley, Reading, 1996.
- [6] Luo Wen. *Reliability Characterization and Prediction of High-K Dielectric Thin Films*. PhD thesis, Texas A& M University, 2004.
- [7] A. Coblenz and L. Harry. *Transistors: Theory and Applications*. McGraw-Hill, New York, 1995.
- [8] G. E. Moore. Cramming more components onto integrated circuits. *IIE Transactions*, 38:1–4, 1975.
- [9] M. Bohr. Intel’s 90nm technology:moore’s law and more. *Intel Developer Forum, Intel Corporation*, 38, 2002.
- [10] G. Siva and N.A. Chandrakasan. *Leakage in Nanometer SiO₂ CMOS Technologies*. Springer, New York, 2006.
- [11] M.J. Deen, W. D. Brown, K. B. Sundaram, and S.I. Raider. *Silicon Nitride and Silicon Dioxide Thin Insulating Films*. Electrochemical Society Inc., New Jersey, 1997.
- [12] Y.E. Wu, J. Sune, and R.P. Vollertsen. Comprehensive physics-based breakdown model for reliability assessment of oxides with thickness ranging from 1 nm to 12 nm. *IEEE 47th Annual International Reliability Physics Symposium*, 47:708–717, 2009.

- [13] R. Degraeve, B. Kaczer, and G. Groeseneken. Reliability: a possible showstopper for oxide thickness scaling. *Semiconductor Science Technology*, 15:436–444, 2000.
- [14] J. Sune, M. Nafria, E. Miranda, X Oriols, R. Rodriguez, and X. Aymerich. Failure physics of ultra-thin SiO₂ gate oxides near their scaling limit. *Semiconductor Science Technology*, 15:445–454, 2000.
- [15] D. J. Dumin. Oxide wearout, breakdown and reliability. *International Journal of High Speed Electronics and Systems*, 11:617–718, 2001.
- [16] M.H. Juang, S. H. Cheng, and Jang S.L. Formation of polycrystalline-Si thin-film-transistors with a retrograde channel doping profile. *Solid-State Electronics*, 53:371375, 2009.
- [17] R.S. Chau. Intels breakthrough in high-k gate dielectric drives moores law well into the future. *Nanotechnology @ Intel Magazine*, 38:1–7, 2004.
- [18] B. Stegemann, S. Daniel, L. Thomas, A. Schoepke, I. Iris, Didschuns, B. Rech, and M. Schmidt. Ultrathin SiO₂ layers on Si(111): preparation, interface gap states and the influence of passivation. *Journal of Nanotechnology*, 19:1405–1409, 2008.
- [19] Y.Z. Yue, Y. Hao, J.C. Zhang, Q. Feng, J.Y. Ni, and X. H. Ma. A study on Al₂O₃ passivation in gan mos-hemt by pulsed stress. *Journal of Chinese Physics*, B 17:1405–1409, 2008.
- [20] Z.Y. Xia, P.G. Han, J. Xu, D. Y. Chen, D. Y. Wei, Z. Y. Ma, L. Chen, K.J.and Xu, and X.F. Huang. Hydrogen passivation effect on enhanced luminescence from nanocrystalline Si/SiO₂ multilayers. *Chinese Physics Letters*, 41:1405–1409, 2007.
- [21] P. Deak, J. Knaup, C. Thill, T. Frauenheim, T. Hornos, and A. Gali. The mechanism of defect creation and passivation at the SiC/SiO₂ interface. *Journal of Physics: Applied Physics*, 41:1405–1409, 2008.
- [22] S. Mahapatra, K.P. Bharath, and M. A. Alam. A new observation of enhanced bias temperature instability in thin gate oxide p-MOSFETs. *IEEE IEDM Technical Digest*, page 337, 2003.
- [23] M. Alam, B. Weir, Silverman P. Bude, J., and A. Ghetti. A computational model for oxide breakdown:theory and experiments. *Microelectronic Engineering*, 59:137–147, 2001.
- [24] E.A. Amerasekera. *Failure Mechanisms in Semiconductor Devices*. Wiley and Sons, New York, 1998.

- [25] J. Sune, D. Jimenez, and E. Miranda. Breakdown modes and breakdown statistics of ultrathin SiO₂ gate oxides. *International Journal of High Speed Electronics and Systems*, 11:789–848, 2001.
- [26] R. Degraeve, G. Groesenken, R. Bellens, J.L. Ogier, M. Depas, P.J. Roussel, and H.E. Maes. New insights in the relation between electron trap generation and the statistical properties of oxide breakdown. *IEEE Transactions on Electron Devices*, 45:904–911, 1998.
- [27] J.W. McPherson and R.B. Khamankar. Molecular model for intrinsic time-dependent dielectric breakdown in SiO₂ dielectrics and the reliability implications for hyper-thin gate oxides. *Semiconductor Science Technology*, 15:462–470, 2000.
- [28] B.P Linder and J. H. Stathis. Statistics of progressive breakdown in ultra-thin oxides. *Microelectronic Engineering*, 72:24–28, 2004.
- [29] R. Degraeve, B. Govoreanu, B. Kaczer, J. Van Houdt, and G. Groesenken. Measurement and statistical analysis of single-trap current-voltage characteristics in ultrathin SiON. In *Proceedings of the International Reliability Physics Symposium*, pages 360–365, May 2005.
- [30] C. O. Chui, F. Ito, and K. C. Saraswat. Nanoscale germanium mos dielectrics - part i: Germanium oxynitrides. *IEEE Transactions on Electronic Development.*, 53:1501–1508, 2006.
- [31] Y. Liangchun, G. Dunne, Matocha K., P. K. Cheung, J. Suehle, and K. Sheng. Reliability issues of SiC mosfets: A technology for high temperature environments. unpublished manuscript: National Institute of Standards and Technology, 16 pages, February 2010.
- [32] K. P. Cheung. Thin gate oxide reliability the current status. unpublished manuscript: Agere System, April 2001.
- [33] I. Polavarapu and G. Okogbaa. An interval estimate of mean-time-to-failure for a product with reciprocal Weibull degradation failure rate. In *Proceedings of the Reliability and Maintainability Symposium*, pages 1–16, Alexandria, VA, Jan 2005.
- [34] W. Nelson. *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*. Wiley, John and Sons, New Jersey, second edition, 2006.
- [35] I. Polavarapu. *Optimal Design of an Accelerated Degradation Experiment with Reciprocal Weibull Degradation Rate*. PhD thesis, University of South Florida, 2004.

- [36] H.F. Yu and C.H. Chiao. Designing an accelerated degradation experiment by optimizing the interval estimation of the mean-time-to-failure. *Journal of the Chinese Institute of Industrial Engineers*, 5:23–33, 2002.
- [37] H.F. Yu and C.H. Chiao. Designing an accelerated degradation experiment by optimizing the estimation of the percentile. *Quality And Reliability Engineering International*, 19:197–214, 2003.
- [38] H.F. Yu. Designing an accelerated degradation experiment with a reciprocal weibull degradation rate. *Journal of Statistical Planning and Inference*, 136:282–297, 2006.
- [39] W.Q. Meeker and LA. Escobar. A review of research and current issues in accelerated testing. *International Statistical Review*, 61:147–168, 1993.
- [40] E. A. Elsayed and L. Jiao. Optimal design of proportional hazards based accelerated life testing plans. *International Journal of Materials and Product Technology*, 17:411–424, 2002.
- [41] H.T. Liao and Z.J. Li. Multiobjective design of equivalent accelerated life testing plans. *Quality and Safety Engineering*, 15:515–538, 2008.
- [42] H.T. Liao. Optimal design of accelerated life testing plans for periodical replacement with penalty. *Naval Research Logistics*, 56:19–32, 2009.
- [43] W.A. Strong. Introduction. In W.A. Strong, Y.E. Wu, P.R. Vollertsen, J. Sune, G. Rosa, S.E. Rauch, and D. T. Sullivan, editors, *Reliability Wearout Mechanisms in Advanced CMOS Technologies*, pages 1–67. John Wiley and Sons, New York, 2009.
- [44] J.F. Lawless. *Statistical Models and Methods for Lifetime Data*. Wiley and Sons, Ottawa, second edition, 2003.
- [45] P.A. Tobias and D. Trindade. *Applied Reliability*. Van Nostrand Reinhold, New York, 1986.
- [46] M.C. Roco. Nanotechnology at NSF: nanoscale science and engineering grantees conference. unpublished manuscript: National Science Foundation, December 2009.
- [47] C.M. Roco. National nanotechnology investment in the fy 2010 budget. unpublished manuscript: National Science Foundation, December 2009.
- [48] E.E. Lewis. *Introduction to Reliability Engineering*. Wiley, John and Sons, New York, second edition, 1996.

- [49] M. Rausand and A. Hoyland. *System Reliability Theory Models, Statistical Methods and Applications*. Wiley, John and Sons, New York, second edition, 2004.
- [50] B. M. Wunderle. Progress in reliability research in the micro and nano region. *Microelectronics Reliability*, 46:1685-1694, 2008.
- [51] Y. C. Yeo, Q. Lu, and C. Hu. MOSFET gate oxide reliability: Anode hole injection and its applications. *International Journal of High Speed Electronics and Systems*, 11:849–886, 2001.
- [52] J. W. McPherson. Physics and chemistry of intrinsic time-dependent dielectric breakdown in SiO₂ dielectrics. *International Journal of High Speed Electronics and Systems*, 11:751–787, 2001.
- [53] J. Sune. New physics-based analytic approach to the thin-oxide breakdown statistics. *IEEE Electron Device Letters*, 22:269–298, 2001.
- [54] J. Sune and Y.E. Wu. Modeling the breakdown and breakdown statistics of ultra-thin SiO₂ gate oxides. *Microelectronic Engineering*, 59:149–153, 2001.
- [55] E.A. Elsayed, H. Liao, and X. Wang. An extended linear hazard regression model with application to time-dependent dielectric breakdown of thermal oxides. *IIE Transactions*, 38:329–340, 2006.
- [56] R. Degraeve, G. Groesenken, R. Bellens, M. Depas, and H.E. Maes. A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides. In *Technical Digest-International Electronic Device Meeting*, volume 45, pages 863–866, 1995.
- [57] J. H. Stathis. Percolation models for gate oxide breakdown. *Journal of Applied Physics*, 86:5757–5766, 1999.
- [58] A.T. Krishnan and E. P. Nicollian. Analytic extension of the cell-based oxide breakdown model to full percolation and its implications. In *Proceedings of the International Reliability Physics Symposium*, pages 232–239, May 2007.
- [59] W. Weibull. A statistical theory of the strength of materials. *Ingeniors Vetenskaps Akademien, Handlingar*, 151-3:45–55, 1939.
- [60] J.B. Watchman, W. R. Cannon, and J.M. Matthewson. *Mechanical Properties of Ceramics*. Wiley and Sons, New York, 2009.
- [61] E. Y. Wu and J. Sune. Power-law voltage acceleration: a key element for ultra-thin gate oxide reliability. *Microelectronic Reliability*, 45:1809–1834, 2005.

- [62] T. Nigam, A. Kerber, and P. Peumans. Accurate model for time-dependent dielectric breakdown of high-k metal gate stacks. In *Proceedings of the International Reliability Physics Symposium*, pages 523–530, Montreal, Canada, April 2009.
- [63] J. Sune, S. Tous, and E. Wu. Analytical cell-based model for the breakdown statistics of multilayer insulator stacks. *IEEE Electron Device Letters*, 30:1359–1361, 2009.
- [64] S. Tous, E. Wu, and J. Sune. A compact analytical model for the breakdown distribution of gate stack dielectrics. In *Proceedings of the International Reliability Physics Symposium*, pages 1–7, Anaheim, CA, May 2010.
- [65] J. Sune and E. Wu. Hydrogen-release mechanisms in the breakdown of thin SiO₂ films. *Physics Review Letters*, 92:1–4, 2004.
- [66] J. Sune, E. Wu, and S. Tous. A physical-based deconstruction of the percolation model of oxide breakdown. *Microelectronic Engineering*, 84:1917–1920, 2007.
- [67] J. Sune and Y.E. Wu. Mechanism of hydrogen release in the breakdown of SiO₂ gate oxides. *Digest of the International Electron Device Meeting*, 1:388–391, 2005.
- [68] Sam Efromavich. *Nonparametric Curve Estimation: Methods, Theory, and Applications*. Springer, New York, 1999.
- [69] Kejian Liu. *Kernel Estimates of Statistical Distribution Functions*. PhD thesis, University of South Florida, 1998.
- [70] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, New York, 2001.
- [71] B.W. Silverman. *Density Estimations for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- [72] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.
- [73] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis*. Springer, New York, 2006.
- [74] Amanuel Teweldemedhin. *Nonparametric Estimators of Mean Residual Life Functions*. PhD thesis, Southern Illinois University, 2008.
- [75] Douglas Montgomery. *Applied Probability and Statistics for Engineers*. Wiley, New York, 2007.

- [76] W. Hardle, M. Muller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer, New York, 2004.
- [77] C. Kuruwita. *A Bayesian Approach for Bandwidth Selection in Kernel Density Estimation with Censored Data*. PhD thesis, Clemson University, 2006.
- [78] An Jia and William Schucany. Recursive partitioning for kernel smoothers: A tree-based approach for estimating variable bandwidths in local linear regression. unpublished manuscript: Southern Methodist University, 31 pages, December 2004.
- [79] B. Miladinovic. *Kernel Density Estimation of Reliability with Applications to Extreme Value Distribution*. PhD thesis, University of South Florida, 2008.
- [80] B. Miladinovic and C.P. Tsokos. Sensitivity of the bayesian reliability estimates for the modified gumbel failure model. *International Journal of Reliability and Safety Engineering, IJRQS*, pages 331–341, 2008.
- [81] M.C. Jones, J. Steve Marron, and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91:401–407, 1996.
- [82] A.W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71:353–360, 1984.
- [83] D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82:1131–1146, 1987.
- [84] M.C. Jones and S.J. Sheather. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, pages 683–690, 1991.
- [85] M.C. Jones, J.S. Marron, and S.J. Sheather. Progress in data-based bandwidth selection. *Computational Statistics*, 11:337–381, 1996.
- [86] H. Shimazaki and S. Shinomoto. A method for selecting the bin size of a time histogram. *Neural Computation*, 19:1503–1527, 2007.
- [87] D.W. Scott. On optimal and data-based histograms. *Biometrika*, 66:605–610, 1979.
- [88] D. Freedman and P. Diaconis. On the histogram as a density estimator: L2 theory. *Probability Theory and Related Fields*, 57:453–476, 1981.
- [89] C.P. Robert. *The Bayesian Choice: From decision-theoretic foundations to computational implementation*. Springer, New York, second edition, 2007.

- [90] P. Congdon. *Bayesian Statistical Modeling*. Wiley, John and Sons, Ottawa, second edition, 2004.
- [91] J.K. Ghosh, M. Delempady, and T. Samanta. *An Introduction to Bayesian Analysis: Theory and Methods*. Springer, New York, 2006.
- [92] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1985.
- [93] H. Rinne. *The Weibull Distribution: A Handbook*. Chapman and Hall: CRC, Boca Raton, FL, 2008.
- [94] J. Rene Van Dorp and T. A. Mazzuchi. A general bayes weibull inference model for accelerated life testing. *Reliability Engineering and System Safety*, 90:140–147, 2005.
- [95] I.D. De Souza and L.R. Lamberson. Bayesian weibull estimates. *IIE Transactions*, 27:311–320, 1999.
- [96] V. Savchuk and C. Tsokos. *Bayesian Statistical Methods with Applications to Reliability*. World Federation Publishers, Russia, 1996.
- [97] P.B. Carlin and A.T. Louis. *Bayesian Methods for Data Analysis*. CRC Press, New York, third edition, 2008.
- [98] W.M. Bolstad. *Introduction to Bayesian Statistics*. Wiley, John and Sons, New Jersey, 2005.
- [99] R.M. Soland. Bayesian analysis of the weibull process with unknown scale and shape parameters. *IEEE Transactions on Reliability*, 18:181–184, 1969.
- [100] G.C Canavos and C.P. Tsokos. Bayesian estimation of life parameters in the Weibull distribution. *Operations Research*, 21:755–763, 1973.
- [101] R.M. Hill. Applying bayesian methodology with a uniform prior to the single period inventory mode. *Journal of Operation Research*, 98:555–562, 1997.
- [102] M. Houssa, G. Pourtois, M.M. Heyns, and A. Stesmans. Defect generation in high-k gate dielectric stacks under electrical stress: The impact of hydrogen. *Journal of Physics: Condensed Matter*, 17:2075–2088, 2005.
- [103] H.Z. Lui, P. Nee, K. Ko, B.J. Gross, T. Ma, and C. Y. Cheng. Field and temperature acceleration of time dependent dielectric breakdown for reoxidized nitrided and fluorinated oxides. *IEEE Electron Device Letters*, 13:41–43, 1992.
- [104] A. Gelman, J. Carlin, H.S. Stern, and B.D. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

- [105] Gelfand. A.E. Sampling-based approaches to calculating marginal densities. *Journal of American Statistical Association*, 85:398–408, 1990.
- [106] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, second edition, 2004.
- [107] N. Ioannis. *Bayesian Modeling Using WinBUGS*. Wiley, John and Sons, New Jersey, 2009.
- [108] M.H. Chen, Q.M. Shao, and J.G. Ibrahim. *Monte Carlo Methods in Bayesian Computation*. Springer, New York, 2000.
- [109] J.E. Griffin and M.F.J. Steel. Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis*, 27:168–176, 2007.
- [110] A. B. Lawson, W.J. Browne, and V.C. Rodeiro. *Disease Mapping with WinBUGS and MLwiN*. Wiley, John and Sons, West Sussex, England, 2003.
- [111] U. Genschel and W.Q. Meeker. A comparison of maximum likelihood and median rank regression for weibull estimation. unpublished manuscript: Iowa State University, 34 pages, June 2010.

APPENDICES

Appendix A Graphical Results: Nonparametric Normal Kernel Probability Density Estimates

The following are graphical results of normal kernel probability density estimates using actual dielectric [6] at the following dielectric failure test stress levels: 8.1, 7.9, 7.7, 7.5, 7.3 and 7.1 MV/cm. The comparison is made between densities estimated using a constant bandwidth (figures to the left) and variable bandwidth (figures to the right).

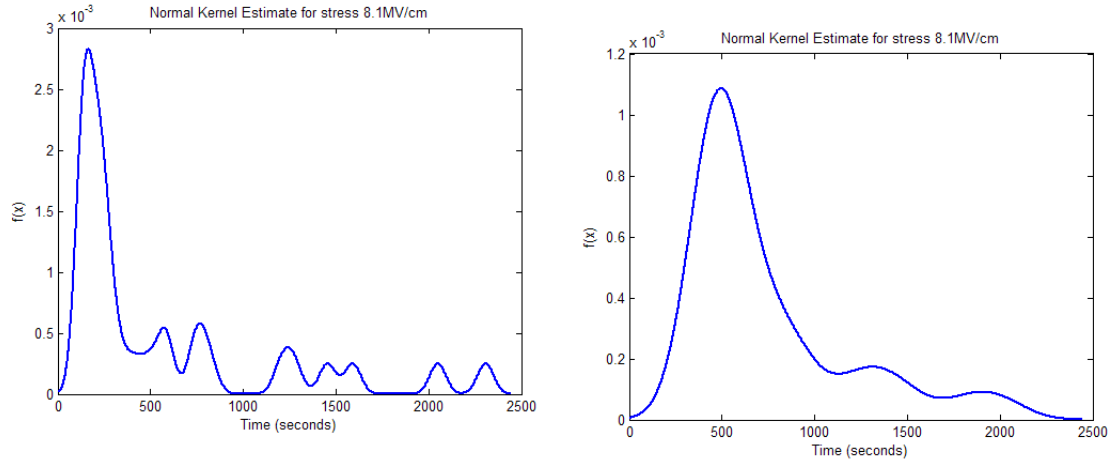


Figure A.1. Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 8.1 MV/cm electric field stress level

Appendix A (Continued)

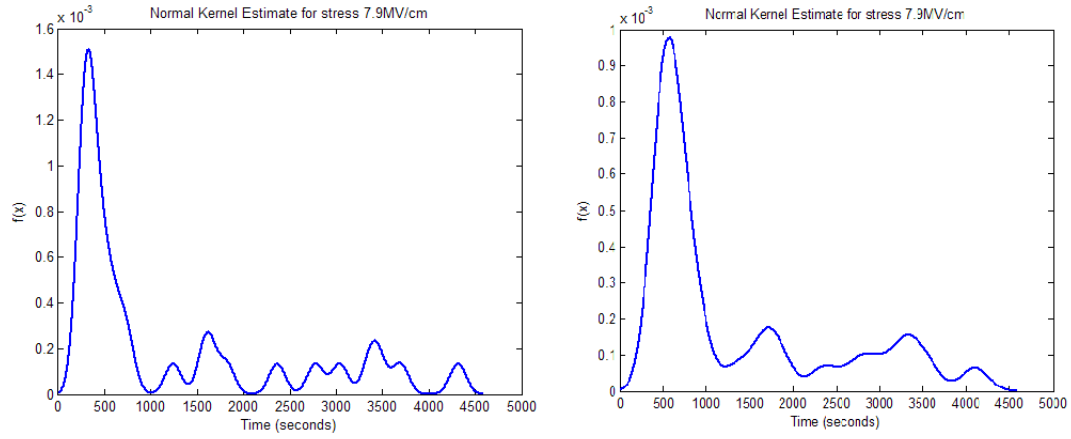


Figure A.2. Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.9 MV/cm electric field stress level

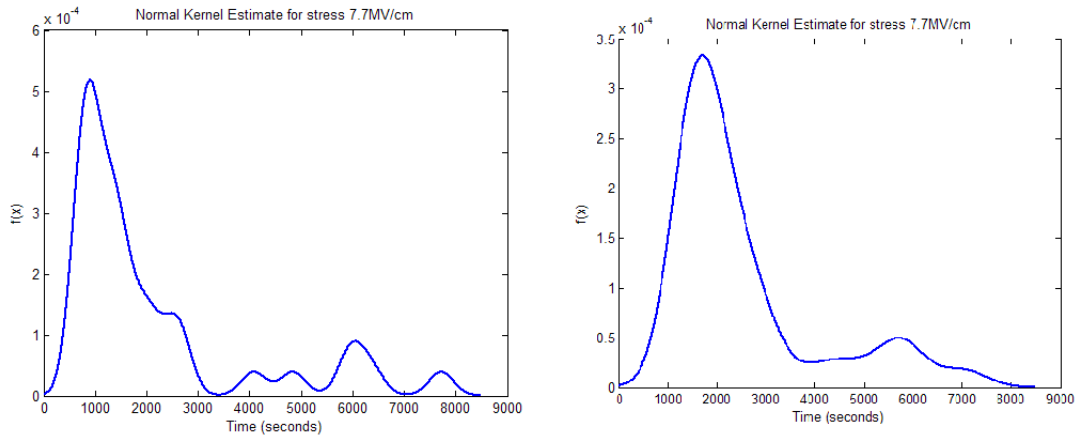


Figure A.3. Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level

Appendix A (Continued)

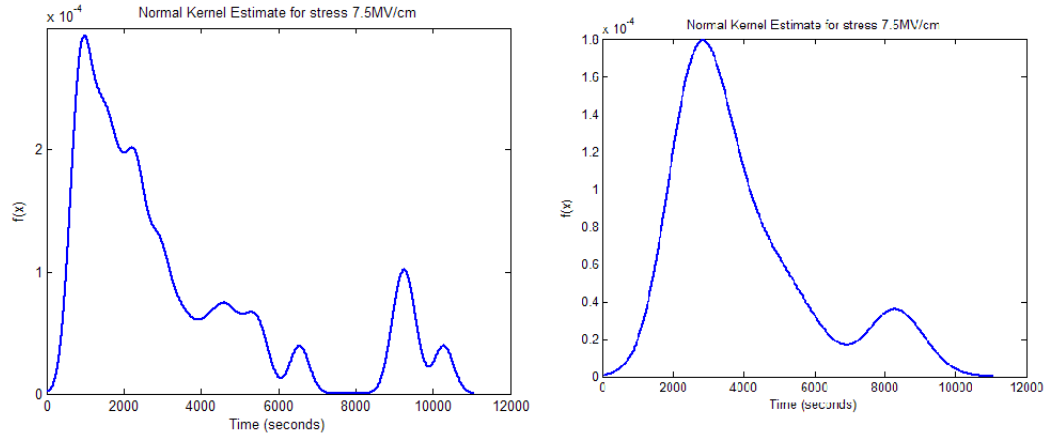


Figure A.4. Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level

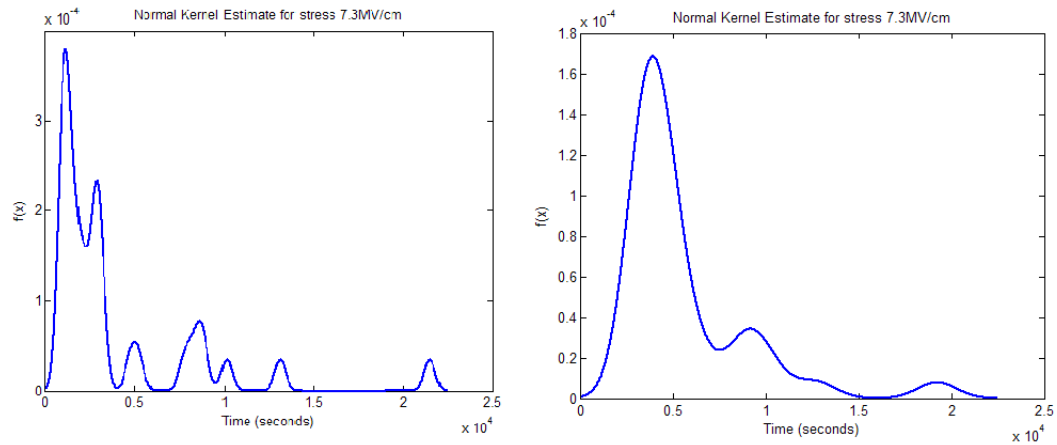


Figure A.5. Normal kernel probability density estimates with constant (right) and variable (left) bandwidth for failure data at 7.7 MV/cm electric field stress level

Appendix A (Continued)

The following are graphical results of Gaussian kernel reliability (figures to the left) and cumulative probability (figures to the right) density estimates using actual dielectric failure data [6] at each dielectric failure test stress levels: 8.1, 7.9, 7.7, 7.5 and 7.3 MV/cm.

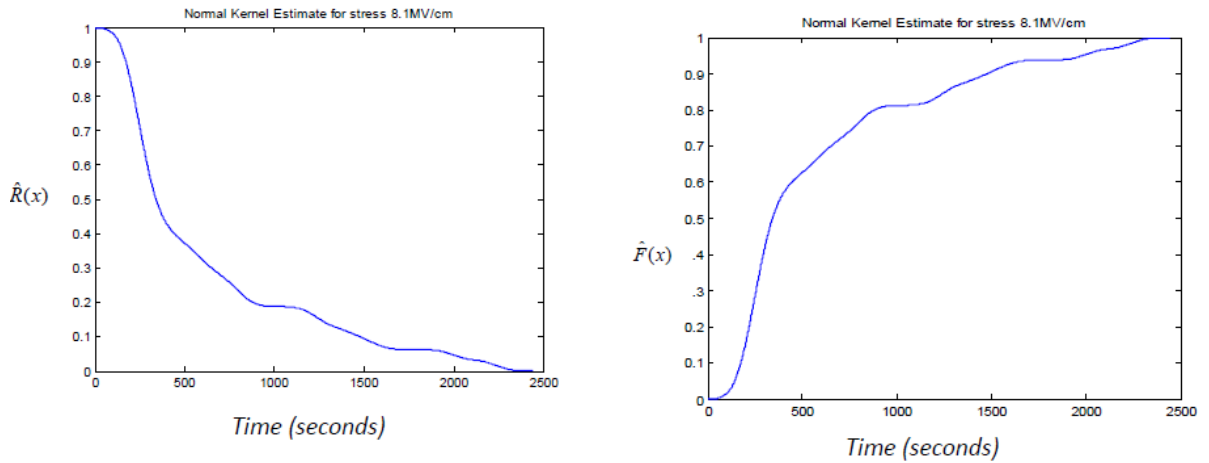


Figure A.6. Reliability (left) and cdf (right) estimates at 8.1 MV/cm electric field stress level

Appendix A (Continued)

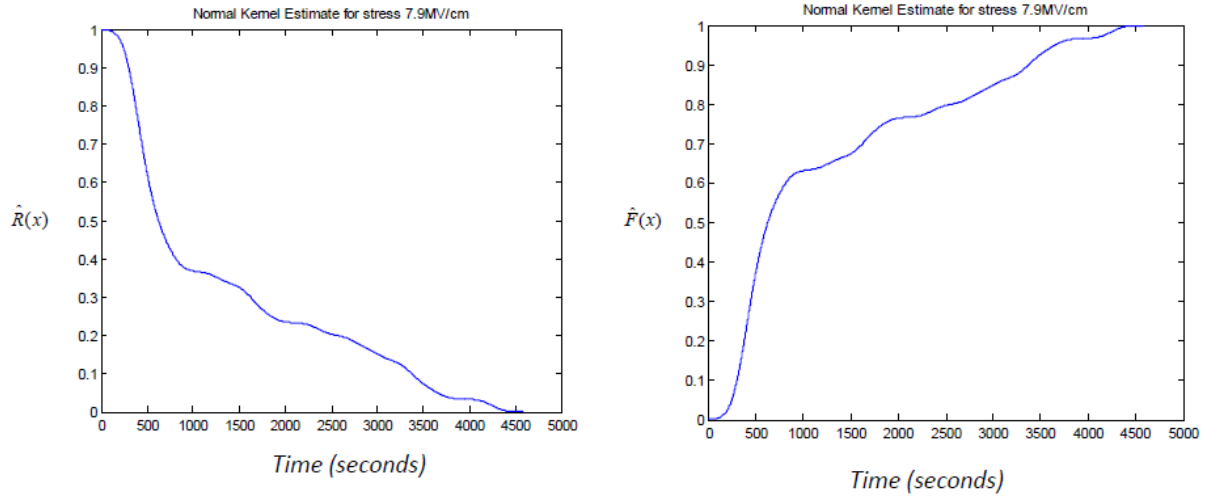


Figure A.7. Reliability (left) and cdf (right) estimates at 7.9 MV/cm electric field stress level

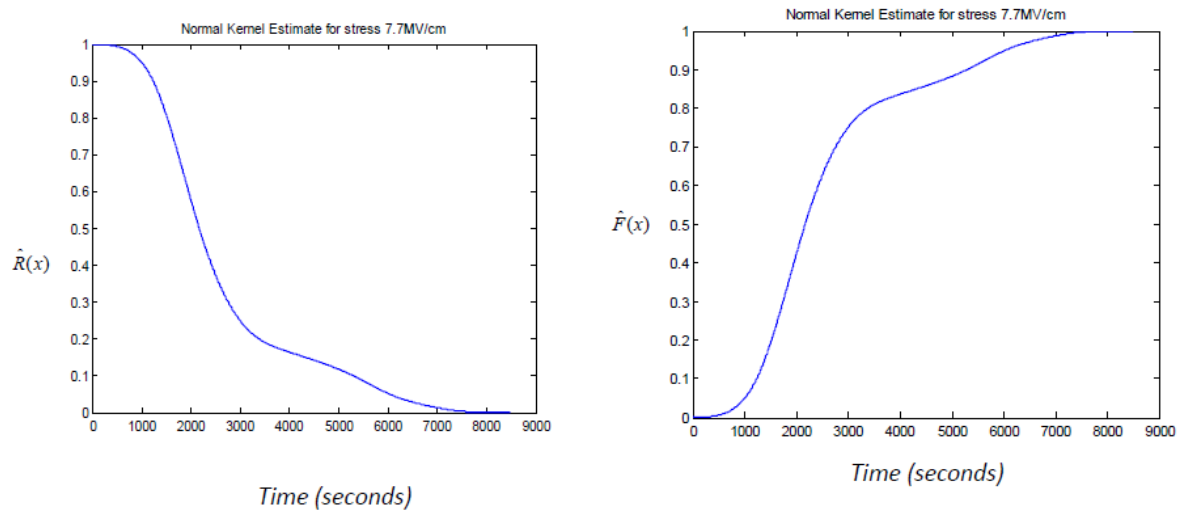


Figure A.8. Reliability (left) and cdf (right) estimates at 7.7 MV/cm electric field stress level

Appendix A (Continued)

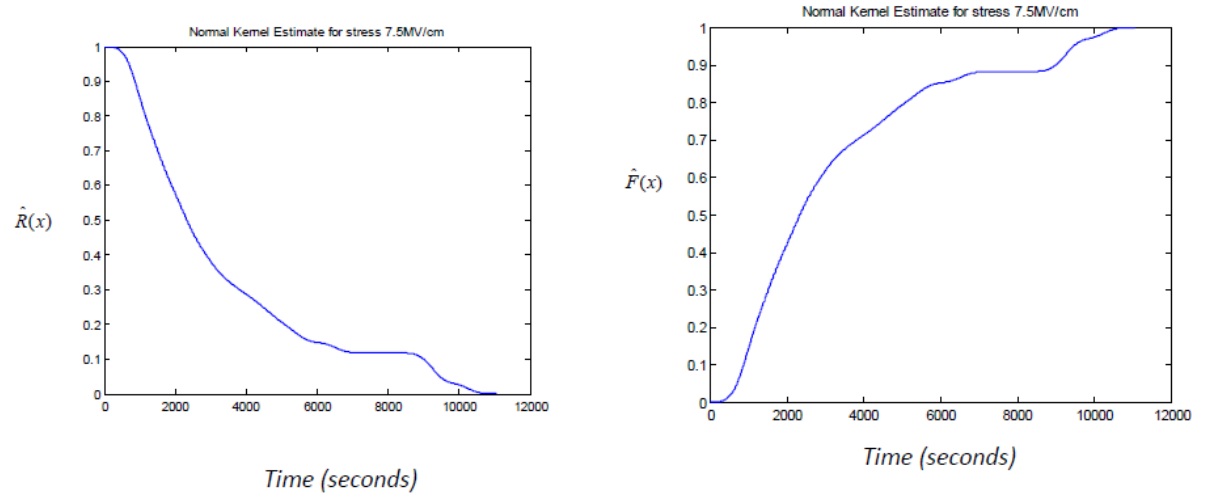


Figure A.9. Reliability (left) and cdf (right) estimates at 7.5 MV/cm electric field stress level

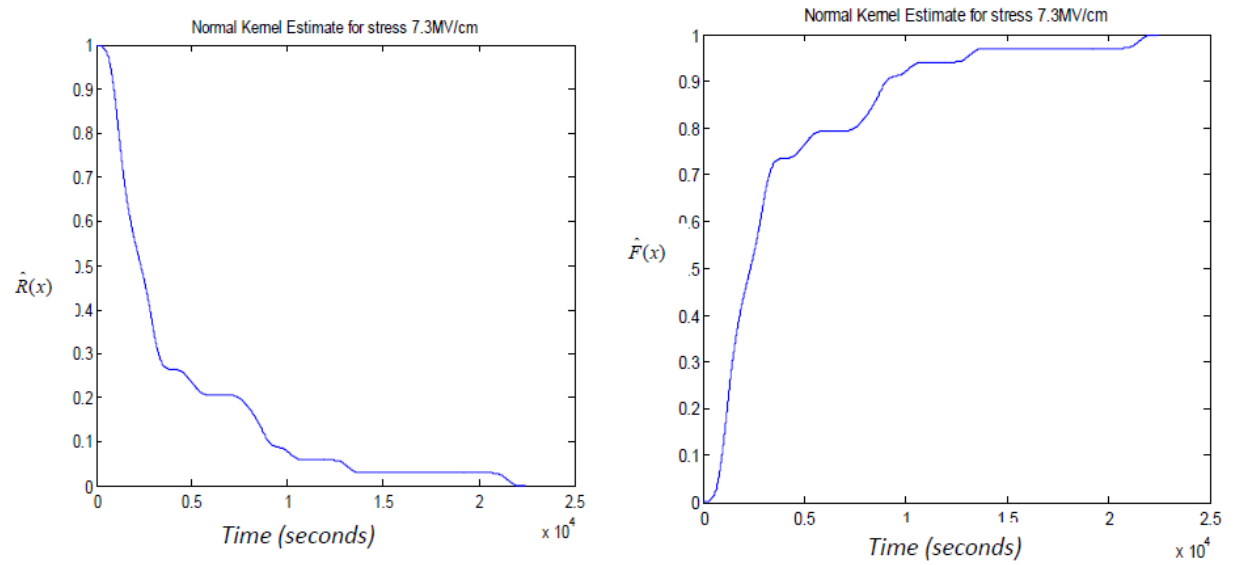


Figure A.10. Reliability (left) and cdf (right) estimates at 7.3 MV/cm electric field stress level

Appendix A (Continued)

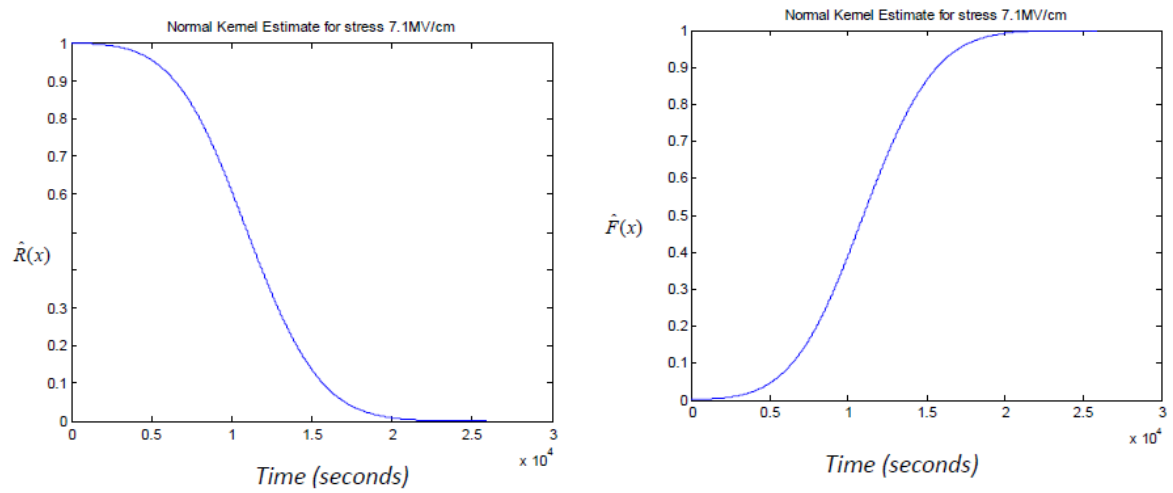


Figure A.11. Reliability (left) and cdf (right) estimates at 7.1 MV/cm electric field stress level

ABOUT THE AUTHOR

Wilkistar Otieno was born in Machakos, Kenya. She earned her Bachelors of Technology Degree with honors in Mechanical Engineering from Moi University in Eldoret, Kenya, where she later worked as a tutorial fellow. She earned her Masters Degrees in Industrial Engineering and Statistics from the University of South Florida (USF) in Tampa, Florida, where she served as an instructor for Probability and Statistics course for Engineers. Her research interests include: reliability of nanoscale materials and devices, survival analysis and design of experiments. Wilkistar served as the manager of a National Science Foundation funded GK-12 STARS program at USF's College of Engineering (2006-2010), and has also been heavily involved with the INFORMS student chapter at USF as vice president, treasurer and editor of their monthly newsletter. She is affiliated with the INFORMS, IIE, TAU BETA PI, ALPHA PI MU and PHI KAPPA PHI professional and honors societies. She has been recognized for her graduate and academic excellence by USF's Graduate School and Phi KAPPA Phi honors society. She was awarded the 2009 Judith Lieberman INFORMS Award for her active role in the INFORMS Student Chapter, and the 2010 USF Graduate School Research Challenge Grant.