RELIABILITY CHARACTERIZATION AND PREDICTION OF HIGH K DIELECTRIC THIN FILM

A Dissertation

by

WEN LUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Industrial Engineering

RELIABILITY CHARACTERIZATION AND PREDICTION OF HIGH K DIELECTRIC THIN FILM

A Dissertation

by

WEN LUO

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Way Kuo (Co-Chair of Committee)	Yue Kuo (Co-Chair of Committee)
Bani K. Mallick (Member)	Yu Ding (Member)
Brett A. Peters (Head of Department)	

December 2004

Major Subject: Industrial Engineering

ABSTRACT

Reliability Characterization and Prediction of
High K Dielectric Thin Film. (December 2004)
Wen Luo, B.E., Tsinghua University;
M.S., University of Washington
Co–Chairs of Advisory Committee: Dr. Way Kuo
Dr. Yue Kuo

As technologies continue advancing, semiconductor devices with dimensions in nanometers have entered all spheres of human life. This research deals with both the statistical aspect of reliability and some electrical aspect of reliability characterization. As an example of nano devices, TaO_x-based high k dielectric thin films are studied on the failure mode identification, accelerated life testing, lifetime projection, and failure rate estimation.

Experiment and analysis on dielectric relaxation and transient current show that the relaxation current of high k dielectrics is distinctive to the trapping/detrapping current of SiO_2; high k films have a lower leakage current but a higher relaxation current than SiO_2. Based on the connection between polarization-relaxation and film integrity demonstrated in ramped voltage stress tests, a new method of breakdown detection is proposed. It monitors relaxation during the test, and uses the disappearing of relaxation current as the signal of a breakdown event.

This research develops a Bayesian approach which is suitable to reliability estimation and prediction of current and future generations of nano devices. It combines the Weibull lifetime distribution with the empirical acceleration relationship, and put the model parameters into a hierarchical Bayesian structure. The value of the Bayesian approach lies in that it can fully utilize available information in modeling uncertainty and provide cogent prediction with limited resources in a reasonable period of time. Markov chain Monte Carlo simulation is used for posterior inference of the reliability projection and for sensitivity analysis over a variety of vague priors.

Time-to-breakdown data collected in the accelerated life tests also are modeled with a bathtub failure rate curve. The decreasing failure rate is estimated with a non-parametric Bayesian approach, and the constant failure rate is estimated with a regular parametric Bayesian approach. This method can provide a fast and reliable estimation of failure rate for burn-in optimization when only a small sample of data is available.

To my dear father and mother

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Way Kuo, for his continuous support and guidance throughout my Ph.D. program. These past five years have been especially rewarding and enjoyable. I learned so many aspects of life from him. I thank Dr. Kuo, from the bottom of my heart, for the great opportunities and help he has given to me. The research topic is challenging and interesting because of his foresight, as well as his insights, in the reliability field. Without his brilliant ideas and enlightening advice, I could not have gone this far.

I am very grateful to my co-chair, Dr. Yue Kuo, for his support on this project and for guiding me into the real world of semiconductor devices. He always provided me light and direction when I was lost in the darkness. It was his patience and encouragement that helped me go through the most difficult time of the research. I feel very fortunate under his guidance, without which I might have given up just a few steps before the finish line.

Dr. Bani Mallick deserves special appreciation for his help and recommendation on Bayesian statistics. I would like to thank Dr. Georgia-Ann Klutke, Dr. Daniel W. Apley, and Dr. Yu Ding for serving on the committee and for giving valuable suggestions while I was working on my dissertation.

I would like to thank Dr. Jun-yen Tewg, Jiang Lu, and Yu Lei of the Thin Film Nano & Microelectronics Research Lab for the sample fabrication and discussions on testing and measurement. Thanks also go to Dr. Kyungmee Kim, Dr. Chunghun Ha, and Dr. Jung Yoon Hwang for the wonderful time and discussions we shared together.

Lastly, I want to thank Liang Jiang, an excellent reliability engineer, who has been a strong and constant cheerleader since the very beginning of this journey.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

As technology has continued to advance and more break-throughs emerge, semiconductor devices with dimensions in nanometers have entered into all spheres of human life [42]. Nowadays, understandings of materials, physics, and reliability at the nano, or atomic, level is vital to the proper design and manufacturing of these devices. It appears that several fundamental barriers restrict the scaling of semiconductor devices. One concern is that tight reliability margins may limit miniaturization. Therefore, there is an urgent need to investigate the reliability of nano devices in the early design stages. A brief description is presented next on some critical issues in the present semiconductor industry. These problems necessitate the research on high k dielectric thin films.

I.1. Reliability concerns with semiconductor devices

The success of the semiconductor industry is largely due to the existence of the gate oxide, SiO_2. A thin film of SiO_2 forms the insulating layer between the control gate and the conducting channel of a metal-oxide-semiconductor field effect transistor (MOSFET). The reliability of SiO_2, i.e., the probability of a SiO_2 film retaining its insulating property under a high electric field for many years, has been the subject of numerous publications.

In the last three decades, the aggressive scaling of microelectronics for better

---

This dissertation follows the style of the *IEEE Transactions on Reliability*.

performance and higher density has maintained an unprecedented pace. Miniaturizing the dimensions and the isolation region of devices is more challenging than ever before as the integration level on a single chip goes to ultra-large scale [63]. It is projected that, by the year 2006, the 0.10μm technology will require an oxide thickness as thin as $1.5 \sim 2.0$nm [24].[1] However, because of performance considerations and non-scalable parameters such as threshold voltage, the scaling of gate voltage lags behind the scaling of oxide thickness. This raises serious concerns as to whether ultra-thin gate oxides can function properly and reliably.

There are reasons to believe that the scaling of oxide thickness is not unlimited. For gate oxide films of less than 3nm, there exists a large leakage current, caused by the direct tunnelling of electrons through the thin layer of SiO_2. This results not only in power loss, but also interferes with the proper function of a MOSFET. Scaling can also be limited by processing issues like poor wafer uniformity, the difficulty of SiO_2 growth control, and the threshold voltage shift caused by boron diffusion from the p^+-gate into the silicon substrate.

In the past, the fabrication of a reliable gate oxide boiled down to sufficiently controlling the processing conditions and lowering the contamination levels to exclude extrinsic failure modes. The intrinsic quality of gate oxides never posed a serious reliability threat under typical operating conditions. But as the oxide thickness decreases, the exponentially increasing tunnelling current can damage the film much faster. Ultra-thin gate oxides may not have sufficient reliability for 10 years of operation [10]. As the intrinsic reliability limits are approached, one solution to the problem is to replace SiO_2 with high k dielectric materials so that the effect of direct tunnelling can be avoided.

---

[1] $1\mu\text{m}=10^{-6}\text{m}$, $1\text{nm}=10^{-9}\text{m}$, $1\text{\AA}=10^{-10}\text{m}$

I.2. MOSFET and MOS capacitor

The schematic structure of a MOSFET is shown in Figure 1. The metal gate electrode is separated from the silicon substrate by the insulating layer of SiO_2. Voltage bias on the gate controls the current flow in the underlying semiconductor region by either creating or eliminating a conducting channel. Working as the switch of a MOSFET, the performance of the gate oxide is crucial to the proper functioning of the transistor. A simple way to test gate oxide is to put it in a metal-oxide-semiconductor (MOS) capacitor, as shown in Figure 2. The metal and the semiconductor correspond to the gate electrode and the silicon substrate of a transistor, respectively. With a gate size of A and a film thickness of d, the capacitance C is calculated as

$$C = \epsilon_o \epsilon_r \frac{A}{d},$$

where ϵ_o is the dielectric constant of the vacuum; and ϵ_r is the relative dielectric constant of a particular material with respect to the vacuum. For SiO_2, $\epsilon_r = 3.9$.

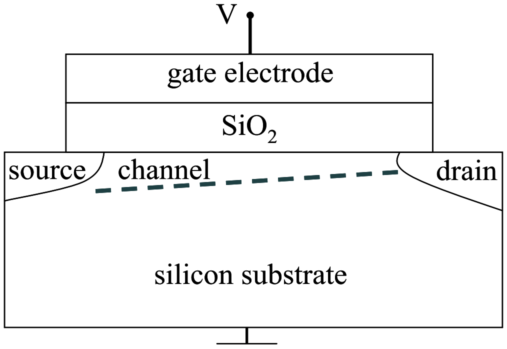


Figure 1: Schematic structure of a MOSFET

A shorter channel is desirable in a transistor for better performance and higher integration [58]. Accordingly, oxide thickness should be reduced for the transistor to

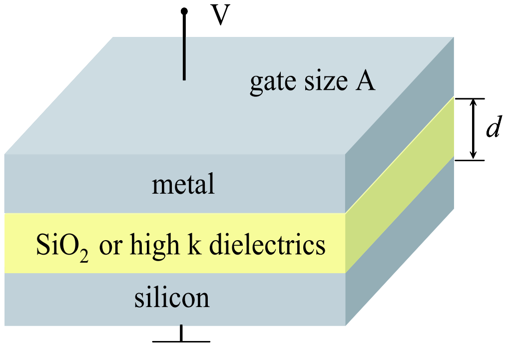


Figure 2: Schematic structure of a MOS capacitor

function properly. This causes problems, such as direct tunnelling and insufficient intrinsic reliability, and limits further device scaling. In order to have a physically thicker film but also maintain the capacitance, the use of dielectric materials with a large value of ϵ_r is the only solution.

I.3. High k dielectric materials

The serious function and reliability problems caused by further reducing oxide thickness can be avoided by replacing SiO_2 with high k dielectric materials, whose relative dielectric constant ϵ_r is in the range of $8 \sim 100$.[2] The benefit of large ϵ_r is that relatively thicker dielectric films can be used without sacrificing film capacitance. Hence, the constraints on scaling set by high tunnelling current, poor film uniformity, and insufficient reliability can be relaxed. For this reason, high k dielectrics are considered as the potential candidates for gate insulators for future generations of transistors [31, 37, 61].

Although high k dielectrics have some deficiencies in material and electric properties, many promising results have been achieved with their use. Difficulties

---

[2] The value of k equals ϵ_r

still exist in merging them into current semiconductor manufacturing technology [61]. The most commonly studied high k dielectrics have been metal oxides, such as tantalum oxide (TaO_x), zirconium oxide (ZrO_x), and hafnium oxide (HfO_x). Ta_2O_5 has high permittivity, while ZrO_2 and HfO_2 are thermodynamically stable on silicon. All are reported to have a much lower leakage current than SiO_2 [39]. There are plenty of encouraging results to warrant further studies on these materials. However, more research is necessary to characterize their reliability properties for a complete assessment of their potential as candidates to replace SiO_2.

Ta_2O_5 is thermally unstable when in direct contact with silicon. A suitable interface layer, such as TaN_y or Si_3N_4, between Ta_2O_5 and the silicon substrate can prevent interfacial reactions and thus improve capacitance and decrease leakage current [29]. A promising strategy for creating better films is the doping of one high k material onto another. A proper dopant can suppress crystallization, decrease leakage and increase the k value. The properties of a hybrid film can be tailored by changing the relative amounts of the individual materials. Ta_2O_5 is especially suited to this technique, because it can be doped with an appreciable amount of another metal oxide to raise the temperature of phase transition from amorphous to polycrystalline [32]. The reliability characterization and evaluation of these TaO_x-based films can provide information of great value to design and fabrication engineers.

I.4. Fabrication of TaO_x-based films

Sample capacitors, with the structure of Al/high k/Si,[3] are fabricated and tested in the Thin Film Nano & Microelectronics Research Lab, Texas A&M University [30]. The high k dielectrics in this study have k values in the range of $10 \sim 15$. As listed in

---

[3] A high k dielectric film between aluminum gate electrode and silicon substrate

Table 1: High k dielectric films under test

High k Dielectric Thin Film		Thickness (nm)	Gate Size (10^{-4}cm^2)	EOT[*] (nm)	k (at 1MHz)
I[†]	TaO_x	6.0	2.27	1.50	NA
	Zr-doped TaO_x	6.8	2.27	2.31	NA
	TaO_x+TaN_y	8.9	2.27	2.06	NA
	Zr-doped TaO_x+TaN_y	9.0	2.27	2.18	NA
II[‡]	TaO_x+TaN_y	10.0	3.72	3.27	11.9
	HfO_x+TaN_y	11.0	3.72	4.24	10.1
	40W Hf-doped TaO_x+TaN_y	9.5	3.72	2.57	14.4
III[♯]	20W Hf-doped TaO_x	62∼63	3.83	2.19	11.3

[*] Equivalent Oxide Thickness.

[†] Used for studying leakage current.

[‡] Used for studying dielectric relaxation and transient current.

[♯] Used for testing time-dependent dielectric breakdown.

Table 1, they are pure TaO_x, pure HfO_x, Hf-doped TaO_x and Zr-doped TaO_x. Some of them have TaN_y as the interfacial layer in contact with the HF pre-cleaned *p*-type silicon wafer, which has an orientation of (100) and a resistivity of 10-80 Ω-cm.

The 5Å thick TaN_y film is sputter deposited on silicon from a Ta target in the Ar/N_2 (1:4 ratio) gas mixture at 5mTorr and room temperature. The Hf-doped (or Zr-doped) TaO_x film, which is co-sputtered from separate Ta and Hf (or Zr) targets, is subsequently deposited in the Ar/O_2 (1:1 ratio) gas mixture without breaking the vacuum. The sputtering power at the Ta target is fixed at 100W, while the power at the Hf (or Zr) target is varied between 20W and 100W, depending on the desired level of dopant concentration. In addition, Pure TaO_x or HfO_x films are deposited at 100W, with or without the TaN_y interface film, to serve as references. The oxide film is then annealed in a quartz tube under O_2 at 600°C for 60 minutes or at 700°C for 10 minutes. The refractive index and thickness of the film is measured with a Rudolph i1000 elliposometer.

The film is fabricated into a MOS capacitor for electrical characterization. The top aluminum electrode is sputter deposited through a shadow mask. The back side of the silicon wafer is also deposited with aluminum for better contact. The post-metal annealing is performed under N_2 at 600°C for 30 minutes. To provide a real sense of the deposited high k thin films, a picture of a 60W Hf-doped TaO_x film taken with a transmission electron microscope is shown in Figure 3 [36].

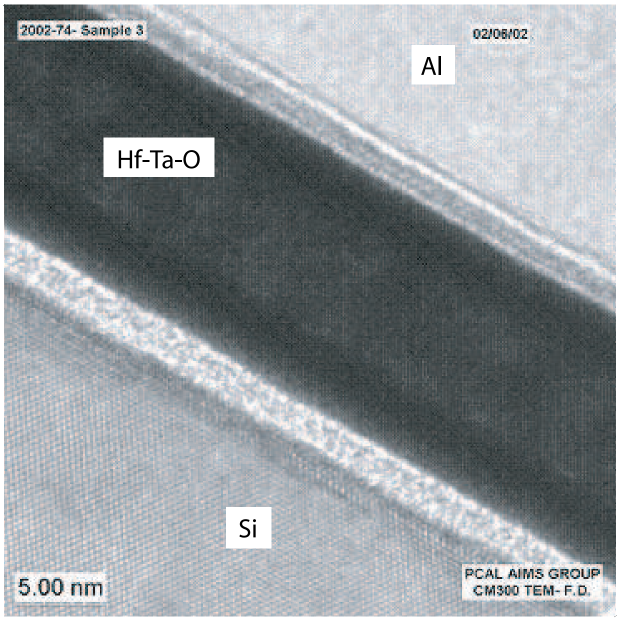


Figure 3: Picture of a 60W Hf-doped TaO_x film taken with a transmission electron microscope [36]

All films are tested at room temperature in a black box to avoid disturbance from light and noise. Figure 4 displays the probe station, hot chuck, and microscope inside the black box. Current versus voltage characteristics are measured with HP 4140B, the pico-ampere meter and DC voltage source. Data of failure times are collected with HP 4155C, the semiconductor parameter analyzer. The output/input signals of the two instruments are transferred in/out of the black box through coaxial and triaxial cables. The HP 4140B and HP 4155C are also connected to a desktop computer for automatic control of the measurements and data acquisition. All tests are programmed in LabVIEW 7.0, a virtual instrument software developed by National Instruments.

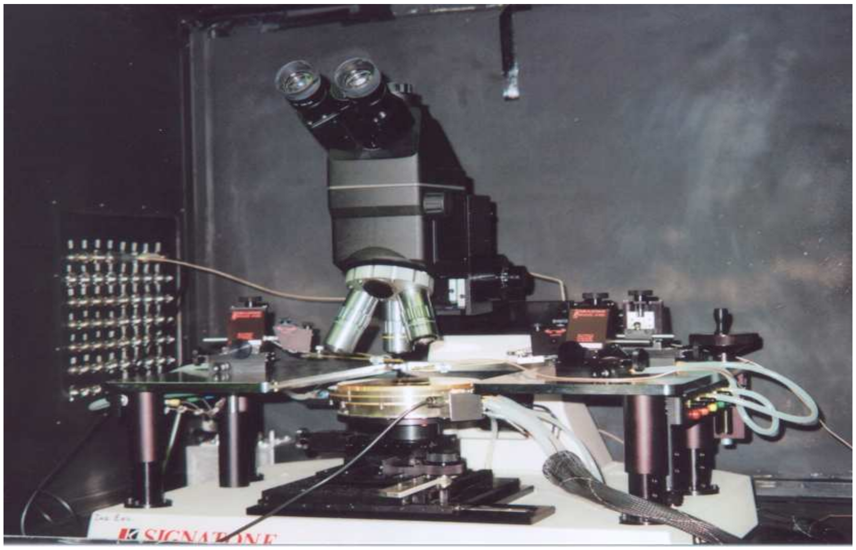


Figure 4: Black box and equipment inside used for testing high k thin films

I.5. Dissertation overview

Chapter II describes several primary reliability characteristics of gate dielectric thin films as well as some commonly used methods of stress testing and breakdown detection. The failure mode of TaO_x-based high k thin films is examined in Chapter III, and a new method of breakdown detection is proposed. Following the experimental study conerning reliability characterization in the previous two chapters, Chapter IV presents a full review of reliability projection for SiO_2 films. A Bayesian approach is then developed in Chapter V to model and project the reliability of high k thin films. Chapter VI continues the topic of the Bayesian approach with discussions on Markov chain Monte Carlo simulation, posterior sensitivity, and inference of lifetime projections. In Chapter VII, the breakdown data from the TaO_x-based high k thin films are used to estimate the failure rates in the first two phases of a bathtub curve, the infant mortality and the useful life. Finally, the reserach is summarized in Chapter VIII.

CHAPTER II

BREAKDOWN AND STRESS TESTS

The reliability of dielectric thin films can be characterized by means of stress testing. While the observed properties strongly depend on the recipe used in fabrication, the test structure and the stressing method also have large impacts on the results of the characterization. To run a stress test, many factors must be carefully considered in advance. Examples of the factors include stress mode, stress time, failure definition, range and resolution of the instruments used for taking measurements, ambient temperature and humidity. It is important to fully understand the strengths and limitations of various stress test methods so that data can be correctly collected, analyzed, and interpreted.

II.1. Reliability characteristics

Reliability is defined as the probability of a device performing its designated function adequately for the intended period of time under the operating condition encountered [28]. For gate dielectric films, the designated function is to act as a sufficient insulator; the operating condition includes mainly the electric field and the temperature that correspond to the normal operation of a transistor; the intended lifetime could be 10 to 20 years, depending on the application [9]. A typical reliability specification for gate dielectric films is formulated as: less than 0.01% of the population with the gate size of 0.1cm^2 is allowed to fail in 10 years under an electric field of 1MV/cm at 100°C.

The main characteristics that describe the reliability of a dielectric film are the voltage to breakdown, V_{bd}; the electric field to breakdown, E_{bd}; the time to breakdown, t_{bd}; and the injected charge to breakdown, Q_{bd} [19]. V_{bd} is the voltage just before breakdown when the gate bias increases gradually, and E_{bd} is the corresponding electric field across the film. If d is the film thickness, then $E_{\mathrm{bd}} = V_{\mathrm{bd}}/d$. E_{bd} is regarded as a gross parameter that represents an estimation of the maximum electric field the dielectric film can sustain, but the film is not subjected to the wear-out kinetics that govern the degradation to breakdown.

Regarding time-dependent dielectric breakdown (TDDB), t_{bd} and Q_{bd} are the quantities that describe the reliability performance [1]. t_{bd} is the time taken to reach breakdown when a stress is applied on the film. It can be measured with two different stress modes, constant voltage stress or constant current stress. Q_{bd} is defined as the cumulative fluence sustained by the dielectric film up until breakdown. It can be obtained by integrating the instantaneous leakage current density, J, over the total stress time, i.e.,

$$Q_{\mathrm{bd}} = \int_0^{t_{bd}} J dt.$$

II.2. Methods of voltage stress testing

Testing methods that have been used to characterize the reliability of gate dielectric films can be categorized according to the stress mode applied to the structure. Commonly used methods include the ramped voltage stress (RVS) test, the constant voltage stress (CVS) test, and the constant current stress (CCS) test [19].

a. Ramped voltage stress test

A RVS test applies a voltage which linearly increases in time on the gate electrode. Very often a staircase voltage, like the one shown in Figure 5, is used instead. Leakage current passing through the dielectric film is monitored at each step and displayed in a current versus voltage (I–V) curve. The occurrence of a breakdown is detected according to a leakage current criterion. The merit of a RVS test is that it is very fast. This type of stress test is used to measure E_{bd}, the maximal electric field a film can sustain before breakdown.

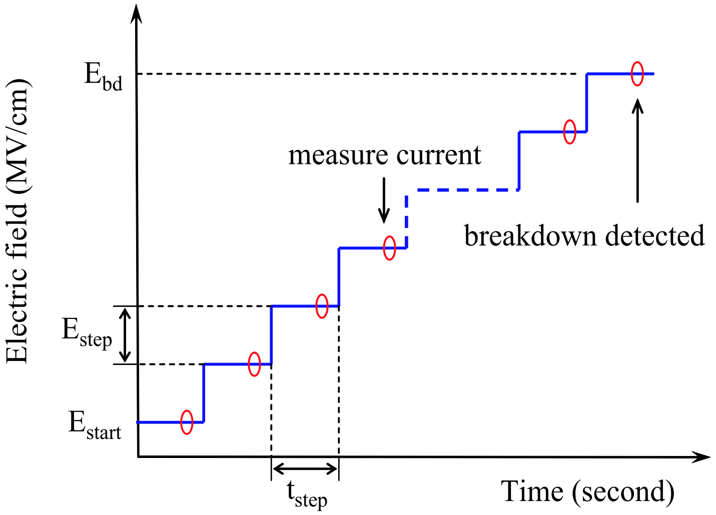


Figure 5: Stress voltage of a staircase RVS

b. Constant voltage stress test

A constant voltage is applied to the test structure while the leakage current flowing through it is plotted against time. This type of stress is more representative of the operating conditions in a real circuit environment than a ramped stress. A CVS test is used to measure t_{bd} and to predict the lifetime under the operating conditions. Breakdown is detected if the leakage current is larger than a threshold or according to a criterion of change in the leakage current.

c. Constant current stress test

In CCS tests, a constant amount of current density is injected into the gate electrode while the voltage evolution over time is monitored through a voltage versus time curve. Breakdown is detected if a sudden drop of voltage occurs between two consecutive stress steps. This test is usually used to measure Q_{bd} and, in some cases, t_{bd} as well.

II.3. Methods of breakdown detection

Breakdown detection is crucial for characterizing the reliability of gate dielectric films, especially as new phenomena, such as stress induced leakage current (SILC) and soft breakdown, demand new interpretations of the measurement. An overestimation of reliability can result if the initial breakdown event is not detected [57]. Due to the wide range of film thicknesses and gate sizes, the multiplicity of test structures, conduction mechanisms, fabrication processes, and test equipments, the idea of a universal failure criterion is not practical. Breakdown criterion must be chosen very carefully for each stress method. Below are some of the commonly used detection methods.

a. Predefined level of leakage current

Thick films should conduct little current up to relatively high electric fields. The measurement of E_{bd} in a RVS test is strongly dependent on the breakdown criterion used. Typically, E_{bd} is the electric field at which the leakage current density exceeds a predefined level [40]. However, if a film is less than 3nm thick, there is significant current conduction at low fields. The main consequence of this is that at high fields, it is difficult to distinguish the tunnelling current during the voltage ramp of the intact dielectric from the current through the resistive path of a film that is badly damaged or destroyed [57]. If a current limit is to be used to define breakdown, then this limit must be determined with reference to some measured conduction and breakdown characteristics of the structure.

b. Increase of leakage current

One way to detect breakdown with a CVS test is to check for a sudden increase of leakage current in consecutive current readings. The relative current change is defined as the ratio of two consecutive currents. However, this ratio is strongly dependent on the film's charge trapping characteristics and measurement settings [40]. This simple definition of breakdown might not be sufficient to identify the first breakdown event and might make it difficult to record both soft and hard breakdowns.

c. Slope increase in I–V curve

To accommodate the testing of ultra-thin dielectric films, more sophisticated and better failure criteria have been implemented. One being used in RVS testing is the slope of the I–V curve [26]. An increase in the slope should be expected at the moment of breakdown.

d. Noise level of leakage current

Soft breakdown in CVS is characterized by a sudden increase in the measurement noise that accompanies the change of leakage current, though the current is well below the level of a hard breakdown [57]. Therefore, an increase in measurement noise can be used to detect soft breakdown in CVS tests [10]. The noise is often determined by calculating the variance of the 10 most recent data points of the leakage current being measured. This method is for CVS tests only and cannot be used with RVS tests.

e. Integrity check

Fowler-Nordheim electron injection due to field-assisted tunnelling is the primary conduction mechanism in SiO_2 films [23]. Since Fowler-Nordheim current is exponentially dependent on the field, the distinction between catastrophic failure and normal Fowler-Nordheim conduction is most clear at low fields where the current differs by several orders of magnitude. Hence, oxide integrity can be examined by means of monitoring the current at a low field [40]. In RVS tests, an integrity check at a low field is performed after each ramp step, as shown in Figure 6. The electric field for the integrity check should be extremely low, around 1MV/cm, and have no significant aging effect on the dielectric. Breakdown is detected if there is a significant increase in the current monitored at the low field.

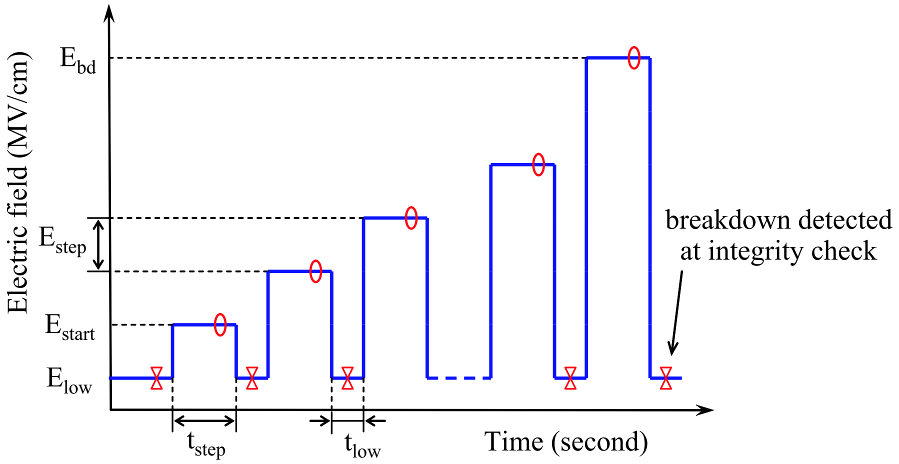


Figure 6: Ramped voltage stress including an integrity check at a low field

II.4. Breakdown mechanisms

Breakdown mechanisms are fairly complex. Although a great deal of effort has been put forth to uncover the physics behind it, there are still many unknowns, debatable issues and controversies. Generally speaking, breakdown is a two-step process, with progressive dielectric degradation followed by final destruction of the film [19]. The degradation is due to the continuous charge build-up that results from the charge injection when the dielectric is subjected to an electric field; the breakdown is initiated by the presence of a weak spot and manifested by the formation of a conductive path through the film.

When a voltage bias is applied to the gate oxide, an electron current will flow if the voltage is high enough and/or the film is thin enough. Depending on the amount of energy, electrons flowing across the oxide will trigger several processes. At least three defect generation mechanisms have been identified [54]. Two of them, anode hole injection and impact ionization, occur at relatively high voltages. They lead to hole trapping and hole-related defect generation. The trap creation attributed to hydrogen released from the anode is the lowest energy process so far identified. As a consequence of the reaction of the released mobile hydrogen, a variety of defects, such as electron traps, interface states etc., gradually build up in the oxide. Eventually, the damage reaches a point where the film breaks down. This defect generation process dominates at the voltage level at which present MOSFETs operate. Other possible breakdown mechanisms also are proposed, but the precise physical details of the trap generation mechanism remain speculative.

Two models of defect generation have been widely discussed in the literature [54]. The anode hole injection (AHI) model claims that breakdown is caused by holes injected from the anode contact. There is a positive feedback mechanism of current runaway, due to the field-enhancement caused by the trapped holes. The AHI model is commonly associated with the time-to-breakdown dependence on the electric field in the form of

$$t_{\mathrm{bd}} \propto \exp(-\frac{a}{E}),$$

where E is the electric field across the oxide and a is a constant.

The thermo-chemical model proposes that defect generation is a field-driven process and that the current flowing through the oxide plays, at most, a secondary role. The model considers the interaction of the applied electric field with the dipole moments associated with oxygen vacancies. The activation energy required for bond breakage is lowered by the dipolar energy. This leads to a quantitative prediction for the field dependence of the activation energy needed for breakdown, and further, the dependence of the time-to-breakdown on the electric field in the form of

$$t_{\mathrm{bd}} \propto \exp(-E).$$

This model has attained widespread acceptance, largely because the breakdown data appears to follow empirically an exponential dependence on the electric field.

CHAPTER III

DIELECTRIC RELAXATION AND BREAKDOWN DETECTION*

The relaxation phenomenon is the recovery of strain on the removal of a stress. The subject of relaxation covers all types of stress relief in solid, dielectric, mechanical, photoconductive, chemical etc. materials. Relaxation has been an intensely researched topic in physics, yet the theoretical understanding of it is still unsatisfactory, largely because of its stochastic nature, insufficient mathematical tools, and experimental data which is very material-related [27].

III.1. Relaxation current

Dielectric relaxation is a bulk-related phenomenon which causes relaxation current following the direction of dV. It has been detected in polycrystalline, disordered, or amorphous films, but not in single-crystal dielectrics [16]. When an external field is applied across a film, it separates the bound charges inside the film. This results in polarization of the dielectric and a compensating internal field [52]. When the external field is released, the internal bound charges are neutralized by the hopping of free charges, but a remnant polarization and an internal field still remain in the film. Because dielectric films have very low conductivity, the hopping process is slow and the relaxation current decays with time following the Curie-von Schweidler law [27]:

---

*©2004 IEEE. Reprinted, with permission, from *IEEE Transactions on Device and Materials Reliability*, "Dielectric relaxation and breakdown detection of doped tantalum oxide high k thin films" by W. Luo, Y. Kuo, and W. Kuo (accepted in August 2004) [38].

$$\frac{J}{P} = at^{-n},$$

where J is the relaxation current density (A/cm^2); P is the total polarization or surface charge density (V $\cdot$ nF/cm^2); t is the time in seconds; a is a constant and n is a real number close to 1.0.

The physical nature of dielectric relaxation can be explained with potential well model in terms of dipole orientation [52]. Dipoles, which are homogeneously distributed inside the material, are formed by localized defects and disorder due to the lack of crystallinity. Dipole orientation is formed as charge carriers hop over potential barriers under the influence of an electric field. However, this charge transfer is confined to a relatively few neighboring sites.

Figure 7 compares the decay of relaxation currents in the tested high k films with those reported in the literature [50]. Relaxation currents are measured after the sudden removal of a constant voltage on the gate. Because the imposed polarization on each film might be different, the relaxation current is normalized to the surface charge density for comparison. The relaxation current is similar in the high k films shown in Figure 7 [16]. However, the relaxation currents of high k films are one to two orders of magnitude larger than that of SiO_2. This observation is consistent with the atomic configuration of high k dielectrics. Firstly, the highly polarizable metal-oxygen bonds can screen the external electric field and, as a result, lead to a large dielectric constant [15]. Secondly, because the metal-oxygen bonds lack symmetry, the center of the positive and negative charges do not coincide after an applied electric field has been removed. Hence dipoles are easily created in high k dielectrics.

Because of its larger magnitude, the relaxation current in high k films can be observed easily. It also has further impacts on the film's behavior. On the one hand, the existence of the relaxation current indicates that the bulk film still maintains

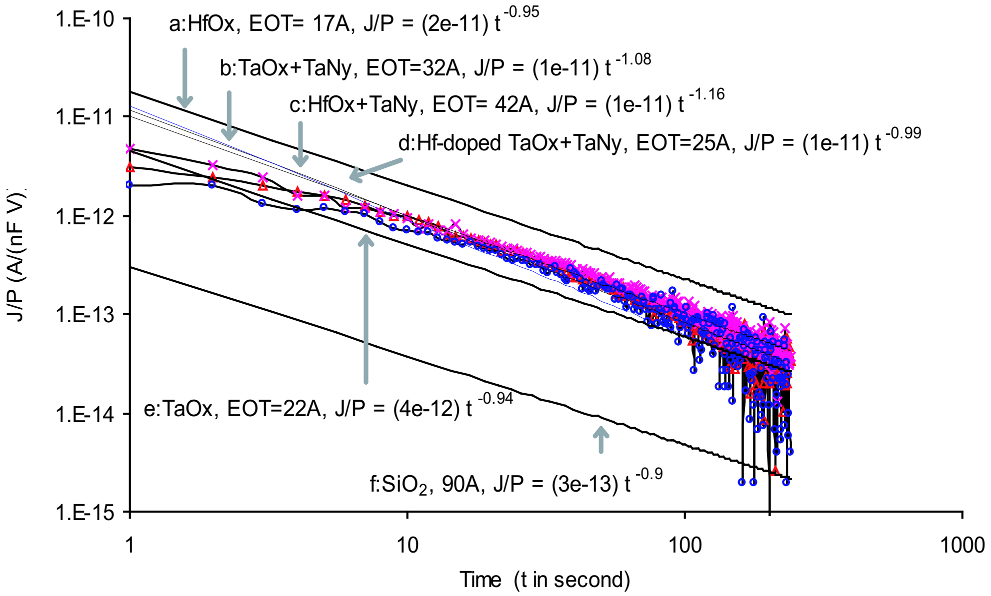


Figure 7: Comparison of relaxation current normalized to polarization

its integrity after being stressed; on the other hand, this transient current can shift flat band voltage and hence severely deteriorate the performance of a MOSFET by changing the threshold voltage and delaying the switching time [16, 50].

III.2. Breakdown mode identification on TaO$_x$-based high k thin films

III.2.1. Static and transient currents

When the voltage bias V changes on a test capacitor, the measured current is a combination of the static current and the transient current. Leakage current has a steady state value, whereas displacement current and relaxation current, both of which depend on dV/dt, are transient. Since the RC time constant of the test capacitors is much less than one second, the observed transient conductivity is contributed mainly by the slow relaxation current [16].

In a strict sense, the polarization-relaxation current originates either from charge separation in the bulk of a neutral film or from redistribution of the space charges that already exist [60]. In the literature, the relaxation current is referred to as the space charge current that corresponds to the buildup of a negative/positive charge in an insulator due to electron trapping/detrapping or the removal of pre-existing space charges. The relationship between the polarization-relaxation current and the space charge current is complicated and is very difficult to separate them. At a low voltage level, the transient current originates mainly from the polarization and relaxation of the dielectric.

In general, static conductivity is contributed by electron transport across the film via a defect band in the bulk, whereas transient conductivity is probably due to dipole orientation, ion motion, or tunnelling into states near the Fermi level [16, 27]. These two conductivities are independent. It is not clear whether they involve the same type of defects [16, 50]. Depending on the polarities of V and dV, the leakage current and the relaxation current, in the strict sense, can have the same or opposite directions. Figure 8 shows the directions of these two currents on a film that is stressed in the accumulation region with a bias that changes from a negative voltage to zero. Under this stress condition and measurement setting, the leakage current is negative and the relaxation current is positive.

III.2.2. Breakdown identification

Although dielectric relaxation is not related to the static leakage current, transient relaxation current can be observed when the leakage current is relatively small. This is demonstrated in Figure 9 with I–V loops measured by sweeping the gate bias from

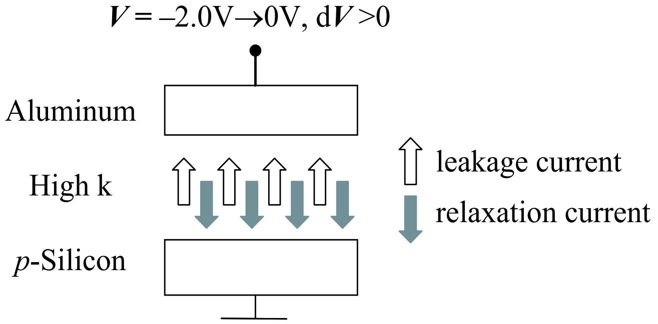


Figure 8: Example of leakage current and relaxation current in a thin film

negative to positive, then back to negative. The hysteresis is obvious on a fresh film in that the I–V characteristics differ as the bias scans in the two opposite directions, i.e., from −2V to +2V and then backward. This demonstrates the existence of transient conductivity in addition to static leakage current. Otherwise, the two I–V curves should be identical regardless of the sweeping direction of the bias voltage. Another piece of evidence is that the measured net current changes direction at the two lowest points of the hysteresis loop. These are the two moments when the magnitude of the leakage current reduces to that of the relaxation current as the two currents flow in opposite directions.

After the first ramp stress that goes from 0 to −5MV/cm on the fresh film, the leakage current increases due to defects generated in the bulk film. As shown in the middle hysteresis loop of Figure 9, the two lowest points at which the measured current changes its direction move closer to 0V than those on the fresh film. The top curve in Figure 9 is measured after the second ramp stress, which is from −5 to −10MV/cm. The dielectric film actually breaks down before −10MV/cm. Because the static leakage current is very high, the hysteresis disappears. It is impossible to determine whether or not transient current still exists at this stage.

Figure 10 shows the relaxation current versus the time since the ramp stress

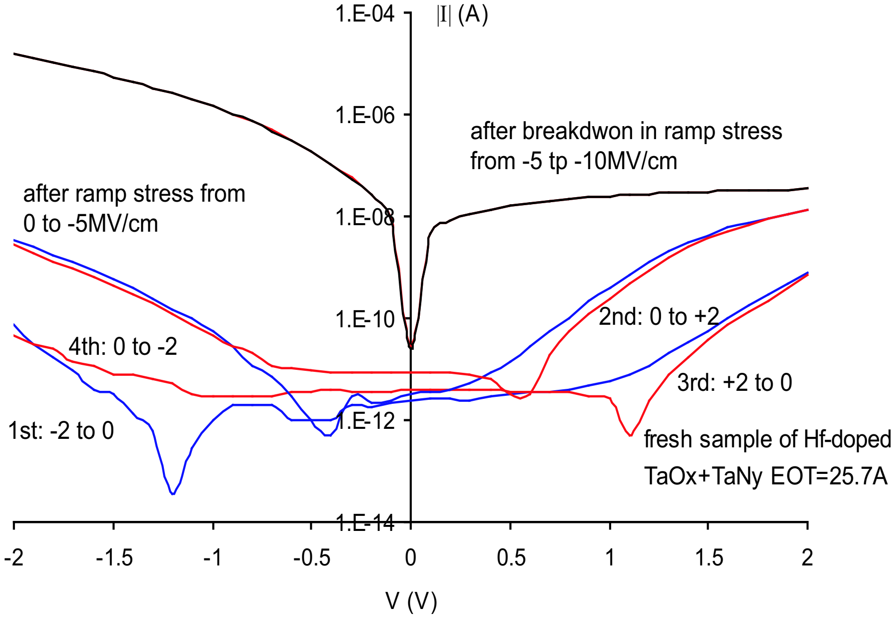


Figure 9: Hysteresis observed in I–V loops

was removed on the same capacitor. The stress and measurement settings are the same as those in Figure 8, except that the gate bias changes from a higher negative bias to zero. The film is still intact after the first ramp stress, and the positive relaxation discharge decreases following the Curie-von Schweidler law. However, after breaking down under the second ramp stress, the film can not regain its strength and shows no sign of relaxation. Because the properties of an insulator are closely related to the relaxation process of the polarized medium [1], an intact film has the resilience to restore its pre-stress state once the stress is released but loses this ability at the moment of breakdown.

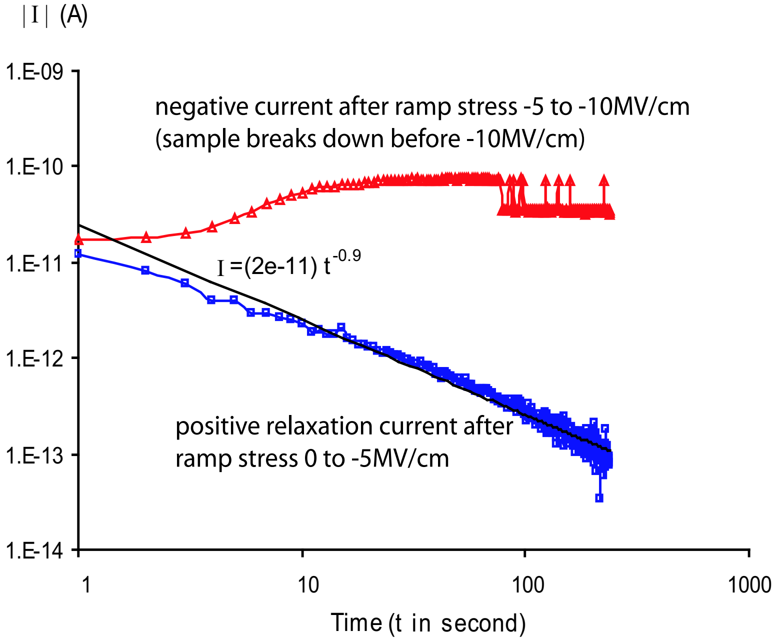


Figure 10: Relaxation discharge before breakdown and disappearance of relaxation current after breakdown

III.3. Transient currents of gate oxides and high k dielectrics

III.3.1. Charging/discharging on gate oxides

Most of the literature on transient current is focused on the charging/discharging of electrons and hole traps [33]. Research on SiO_2 shows that both flat band voltage shift and discharging current are proportional to t^{-1} [4, 13, 47]. Although this relationship is similar to the Curie-von Schweidler law, it is derived from the tunnelling front model, in which the time dependence of the flat band voltage shift after an avalanche injection of holes is modeled by the motion of a tunnelling front moving through the oxide [13]. When high-voltage stressing generates a large number of traps uniformly distributed within the oxide, the current decay due to these traps, after removal of the stress, is proportional to t^{-1}, whereas the magnitude of the current is proportional to the trap density.

A study of the degradation and breakdown of thin oxides films concludes that background trapping in as-fabricated bulk-oxide sites (due to the presence of water or OH) is not an important contributing factor to the total negative-charge buildup in most cases [12]. The two main degradation mechanisms are electron-trap creation and impact ionization caused by hot carriers at high electric fields. After a certain amount of current injection, electron-trap creation occurs when hot electrons release hydrogen from defect sites near the anode interface. The mobile hydrogen moves to the cathode-oxide interface where it produces interface states and electron-trap distribution nearby. The impact ionization process occurs at fields exceeding 6∼7MV/cm. Mobile holes produced in the oxide bulk closer to the anode move under the applied field to the cathode-oxide interface where some are trapped in energetically deep sites (due to oxygen vacancies). Some of the injected electrons from the cathode recombine with these trapped holes to produce interface states and traps near the cathode. Strong detrapping occurs only at high fields, and the dominant mechanism for this is the tunnelling from the ground state of the trap to the bottom of the conduction band of the oxide.

III.3.2. Root causes of transient current

Although transient currents in SiO_2 and high k dielectrics seem to have similar behaviors, such as decaying with a time dependence of t^{-1}, their main contributions are different. Because SiO_2 does not have a significant polarization effect, the background trapping in as-fabricated bulk films and trap creation at low electric fields are limited. However, a large amount of trapping occurs under high field stresses due to electron-trap creation and impact ionization. This results in trapping/detrapping current, which is often called the relaxation current in SiO_2.

In high k dielectrics, the large relaxation current originates mainly from the highly polarizable metal-oxygen bonds. Dipoles and polarization charges formed under an electric field are the reasons behind the pronounced relaxation behavior [22]. The polarization charges are present in the transition region of the heterogeneous medium, in which local variations of permittivity and conductivity (due to non-stoichiometry, impurity, vacancy, dipole, and etc.) exist. They form potential wells which can trap charge carriers. The latter can be locally detrapped by the Schottky effect to produce transient current and dipole orientation. The I–V hysteresis loop of the fresh film in Figure 9 shows that a relatively large relaxation current can be observed in high k dielectrics, even under low stress conditions. This kind of polarization-relaxation behavior is consistent with the general relaxation phenomena defined in physics.

Reports in the literature as well as the experimental data collected in this study, are useful in differentiating the transient current in high k dielectrics from that in SiO_2. However, as mentioned previously, the relationship between the polarization-relaxation current and the space charge current is complicated. More extensive study is required to fully understand the transient conductivity of high k dielectric films.

III.4. Breakdown detection on TaO_x-based films

Identifying the exact moment of a breakdown is the key issue in failure data collection. A breakdown event is usually detected with one of the methods described in Section II.3. The first abrupt jump in leakage current is often considered to be the signal of breakdown. But for a very thin film, due to the fact that a large leakage current exists in the background, this method is misleading when either there is no jump or the jump observed is not due to breakdown. Since no jump is observe on some tested high k films, the method is not recommended. Other methods, such as using a relative increase of leakage current or a slope increase of the I–V curve, are not practical, because it is difficult to specify a breakdown criterion when the measured I–V curves are smooth and increase gradually.

In the process of identifying the breakdown mode of TaO_x-based films, we observed a connection between dielectric relaxation and the breakdown event. Based on this finding, a new detection method is proposed here. This method can be categorized as an integrity checking procedure, using relaxation current instead of Fowler-Nordheim current.

III.4.1. Proposed method of breakdown detection

I–V hysteresis and relaxation discharge shown in Figure 9 and Figure 10, respectively, indicate that the relaxation property is closely related to film integrity. To further investigate the change in relaxation behavior under stress, we designed a RVS test that is depicted in Figure 11. The important feature of the test is that the relaxation current is monitored at a zero gate bias after each ramp step. The purpose of choosing the zero bias as the checking state is to monitor a relatively large relaxation current while the accompanying leakage current is relatively small. The interval at zero bias, t_M, should be at least long enough for the instrument to make an accurate current reading. t_D, the time between the start of t_M to the point at which the relaxation current is measured, can influence the magnitude of the measurement, but it does not change the overall pattern of the data.

Figure 12 shows I–V curves that are measured using the proposed method. The directions of the leakage current and the relaxation current are the same as those

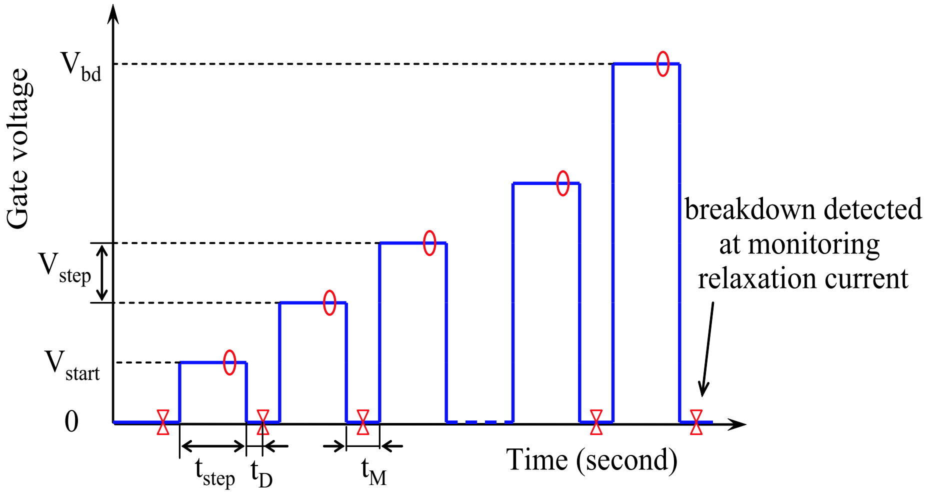


Figure 11: Proposed method of breakdown detection

given in Figure 8. The upper half of the vertical axis represents the magnitude of the negative leakage current, while the lower half represents the magnitude of the positive relaxation current. The further a current is from the horizontal axis, the larger magnitude it has. Note that every current monitored at zero bias has a sudden change of direction at the moment when the relaxation current disappears. More importantly, this coincides with the jump of the leakage current if a jump does occur, as in the cases of the $\mathrm{HfO}_x+\mathrm{TaN}_y$ film and the Hf-doped $\mathrm{TaO}_x+\mathrm{TaN}_y$ film. Since this jump signals a breakdown, the coincidence indicates that the disappearance of relaxation can be used for the same purpose as well. For the $\mathrm{TaO}_x+\mathrm{TaN}_y$ film, although the leakage current does not have an abrupt jump, the relaxation current disappears in the same way as in other tested films. A second measurement of the I–V characteristic confirms that the film did break down from the previous ramp stress test.

Figure 13 compares the two I–V curves of a $\mathrm{TaO}_x+\mathrm{TaN}_y$ film measured with the method proposed. The vertical axis gives the current magnitude, while the direction of each current is described in the figure. The positive relaxation current

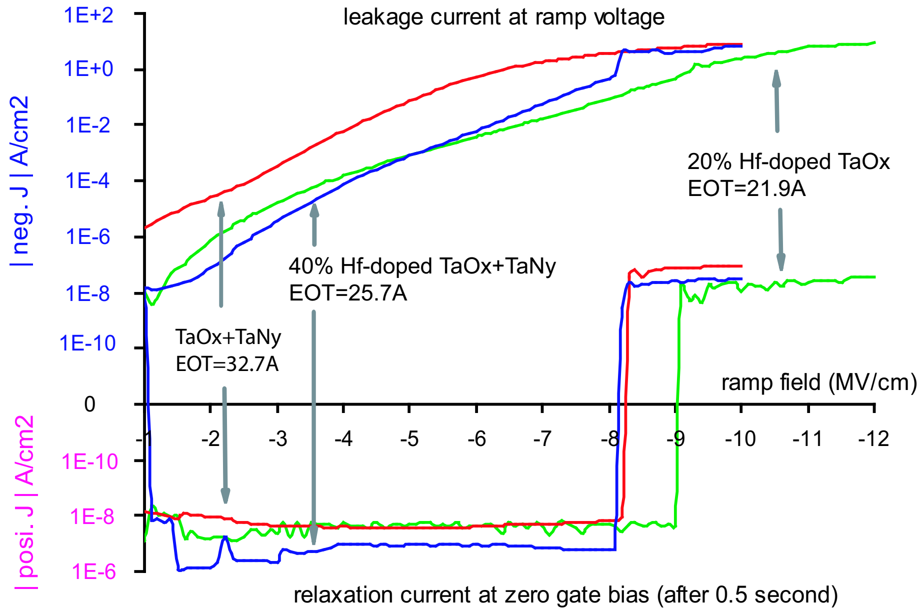


Figure 12: Leakage current and relaxation current versus electric field, measured with the proposed method

monitored at the zero bias disappears when the ramp stress reaches −7.8MV/cm. The fact that the film breaks down at this moment is confirmed by the second I–V measurement, in which the leakage current at low ramp steps increases by more than two orders of magnitude. The reason that the jump of leakage current at breakdown is not observed on this film is because a large leakage current already existed before breakdown.

III.4.2. Remarks on relaxation and breakdown

Many studies have found that high k films have Schottky emission or Poole-Frenkel conduction mechanism rather than Fowler-Nordheim tunnelling, which is the case for SiO_2 [1]. Therefore, the integrity check used for SiO_2 is not suitable for high k films. The breakdown detection method proposed in Figure 11 uses the fact that

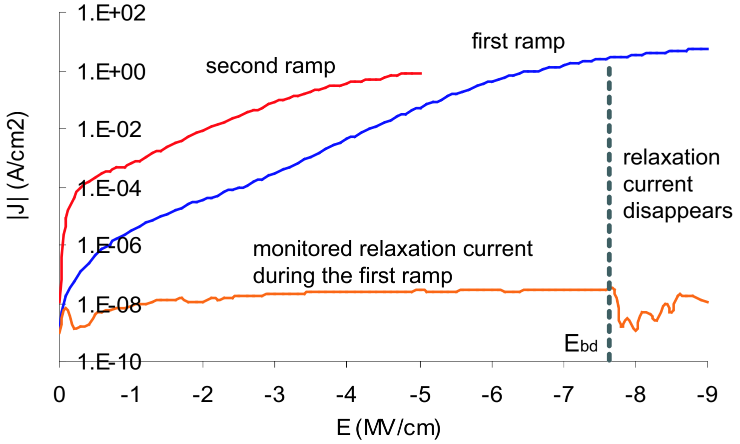


Figure 13: Two consecutive I–V curves measured on a $TaO_x + TaN_y$ film with the proposed method

the relaxation behavior and film integrity are closely related. This is supported not only by the experimental data, but also by the theory that breakdown is due to the relaxation of polarization energy after the trapped charge distribution collapses [22].

It is generally accepted that an oxide film breaks down when the trap density exceeds a critical level [13]. Gressus and Blaise also included the factor of dielectric polarization and relaxation in explaining the breakdown process [22]. When a critical space charge field is built up in a local region, the charges trapped in potential wells are detrapped by the Schottky effect and spread out into surrounding traps until the space charge field becomes equal to, or smaller than, the critical field. If a local detrapping process is strong enough to spread charges to a large area and destabilize the space charge distribution, then destructive breakdown occurs. This involves the propagation of an electric wave associated with the detrapping process and a thermal shock wave which initiates the formation of plasma. As dielectric polarization and relaxation contribute to a large part of the transient current in high k films, Gressus and Blaise's statement supplies a good explanation for why relaxation current can be used as an indicator of film integrity.

The connection between dielectric relaxation and breakdown events on the TaO_x-based films can be easily observed, because they have a higher relaxation current but a lower leakage current than SiO_2. Relaxation currents have been compared in Figure 7, and a comparison of leakage currents is given in Figure 14. The leakage currents of SiO_2 are measured on $Al/SiO_2/Si$ capacitors with the gate size of $1.77 \times 10^{-4}\mathrm{cm}^2$ [3].[1] It is obvious in Figure 14 that, with similar equivalent oxide thickness, the leakage currents of TaO_x-based films are consistently lower than those of the SiO_2 films.

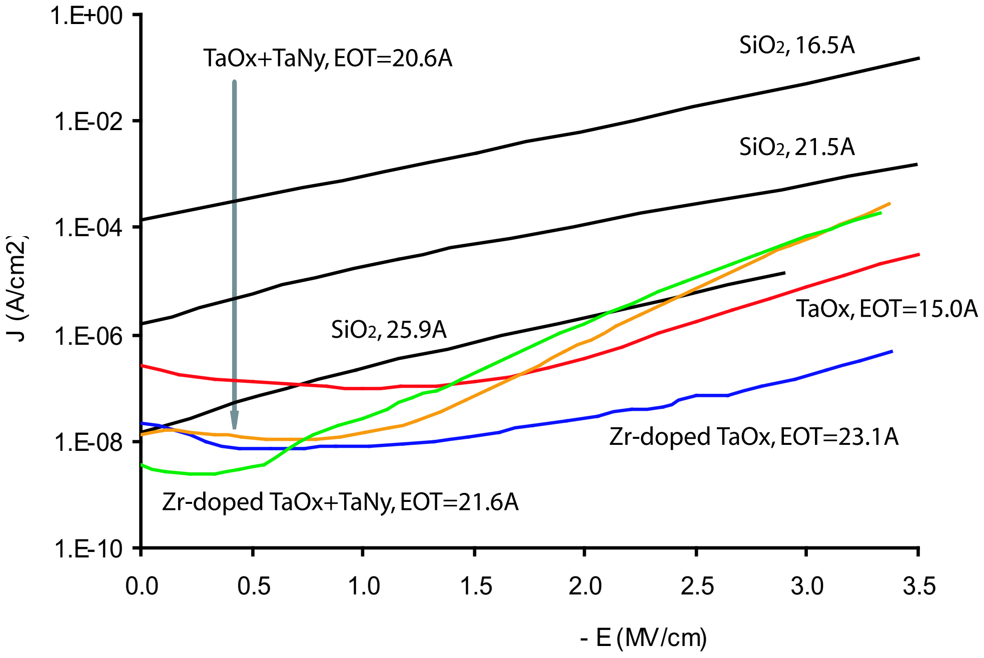


Figure 14: Comparison of leakage current between high k and SiO_2 films

To summarize, the relaxation property is an essential element of film integrity. Because the behavior of polarization and relaxation directly relates to breakdown, the

---

[1] A SiO_2 film between aluminum gate electrode and silicon substrate

sudden disappearance of relaxation current can be used as the signal of a breakdown event. The proposed method of breakdown detection works successfully for identifying breakdowns in the TaO_x-based films. It is very sensitive when films are stressed in accumulation, in which case the relaxation and the leakage currents flow in opposite directions. This method avoids the difficulty of setting an empirical and quantitative criterion as is required by some common detection methods. It is simple and accurate even when an abrupt change in leakage current does not occur. This makes it most useful during RVS tests, in which a high k film fails at a relatively high field and the large leakage current in the background obscures the current jump caused by the breakdown.

CHAPTER IV

RELIABILITY MODELING AND PROJECTION OF GATE OXIDES

Since the first MOSFET was invented in the mid 1970s, a great amount of research has been conducted on the reliability of gate oxide SiO_2. Although there are still issues under debate, large and significant advances have been made in understanding the statistical behaviors, as well as the physical mechanisms, of oxide breakdown. In replacing SiO_2 with high k dielectrics, although the differences in the material properties lead to changes in the electric and reliability characteristics, most of the fundamental theories and techniques are still applicable. Hence, an understanding the statistical and physical models of SiO_2 breakdown provides an advantageous starting point for studying high k thin films.

IV.1. Statistical breakdown model

Breakdown characteristics, such as E_{bd} and t_{bd}, are random variables, each of which follows a certain statistical distribution. There was an early debate about which of the two distributions, Weibull or log-normal, is suitable for t_{bd}, since data sets collected from various stress tests usually fit both equally well. But there is a major difference between the Weibull and log-normal when the fitted model is used to extrapolate the lower percentiles beyond the range of sample data. The debate is resolved through looking for the theoretical justification of a model based on the failure mechanism under investigation.

The precise model that leads to a log-normal distribution is a multiplicative or proportional growth model [59]. At any instant in time, the degradation process undergoes a random increase which is proportional to the present state; the multiplicative effect of all of these random, independent growths of degradation builds up to the failure. Hence a log-normal model is preferred if a multiplicative degradation process is going on. The Weibull distribution is known as the extreme value distribution that describes the smallest of many non-negative random variables. Its derivation suggests its application for a "weakest-link" type of problems. If there are many identical and independent competing processes that lead to the failure of a device, and the first one reaching a critical stage determines the failure of the device, then the lifetime of the device follows a Weibull distribution.

Both experimental data and the physical examination of thin oxides provide direct evidence for the justification of using the Weibull distribution [62]. The true rationale for rejecting the log-normal model is that the breakdown of gate oxide is a weakest-link problem. The failure of a microchip, which consists of millions of transistors, is defined by the first failure occurring on any one of these individual transistors. Similarly, a transistor fails if any small portion of its gate area breaks down [54].

Let X be the random variable that represents t_{bd}. Then X follows a Weibull distribution, denoted as $\mathrm{W}(\alpha, \beta)$, with the cumulative distribution function of

$$F(x) = 1 - \exp[-(\frac{x}{\alpha})^{\beta}], \tag{4.1}$$

where α and β are the scale and shape parameters, respectively. Note that $F(\alpha) = 63.2\%$ at $x = \alpha$. In other words, α equals the 63.2^{th} percentile lifetime. Because of this, α is also called the characteristic life.

After some manipulation, the equation (4.1) can be written as

$$\ln\{-\ln[1 - F(x)]\} = \beta(\ln x - \ln \alpha). \tag{4.2}$$

Then, plotting $\ln\{-\ln[1-F(x)]\}$ versus $\ln x$ results in a straight line with a slope of β and an intercept of $\ln\alpha$, as shown in Figure 15. This plot is usually called a Weibull plot, and β is also called Weibull slope.

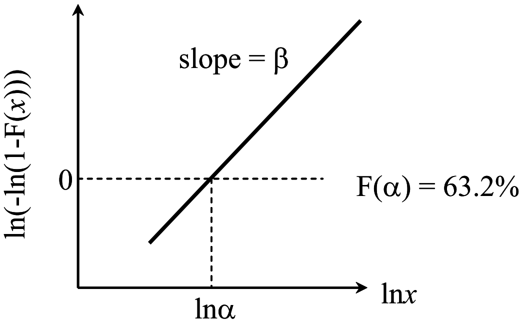


Figure 15: Weibull distribution on Weibull paper

A Weibull distribution has a closure property [59]. If a system is composed of n parts, each having an independent Weibull lifetime distribution with a common β but not necessarily the same α, then the system's lifetime, which is determined by the first part failure, follows a Weibull distribution with the same β.

IV.2. Area scalability

In modeling the reliability of a microchip, area extrapolation from a small test structure to the entire chip area is indispensable. To derive the area dependence of breakdown, consider a capacitor with a top area of A^*, where A^* is much smaller than the smallest capacitor available for testing. Two assumptions are hold on the small capacitor [46]: the breakdown occurs at a random position, independent of the mechanism that causes it; and t_{bd} follows $\mathrm{W}(\alpha, \beta)$.

A large capacitor with a top area of A_1 can be considered to be built up of n such small capacitors, where $n = A_1/A^*$. By the closure property of the Weibull distribution, the time-to-breakdown of the large capacitor follows a Weibull distribution with the same β as the small capacitors. The reliability of the large capacitor at time x is the probability that none of the n small capacitors has failed up to x. That is,

$$1 - F_1(x) = [1 - F^*(x)]^n = \{\exp[-(\frac{x}{\alpha})^\beta]\}^{\frac{A_1}{A^*}},$$

or equivalently,

$$\ln\{-\ln[1 - F_1(x)]\} = \ln\frac{A_1}{A^*} + \beta \cdot \ln\frac{x}{\alpha}. \tag{4.3}$$

For two large capacitors with top areas of A_1 and A_2, if both are built up with the same type of small capacitor A^*, then their Weibull lifetime distributions have the same slope β but different characteristic lives, say α_1 and α_2. Then, equation (4.3) leads to

$$\ln\{-\ln[1 - F_2(x)]\} - \ln\{-\ln[1 - F_1(x)]\} = \ln\frac{A_2}{A_1}. \tag{4.4}$$

This means that if the top area increases from A_1 to A_2, the complete breakdown distribution on a Weibull plot shifts up a vertical distance of $\ln(A_2/A_1)$. Using equation (4.1) to replace F_1 and F_2 in equation (4.4) gives the area dependence of the characteristic life as a function of the Weibull slope β,

$$\frac{\alpha_1}{\alpha_2} = (\frac{A_2}{A_1})^{\frac{1}{\beta}}. \tag{4.5}$$

IV.3. Accelerated life testing

It is unlikely that within a reasonable time and at an acceptable cost, one could directly test whether or not a reliability specification for normal operation is satisfied. First of all, the lifetime under the operating condition could be years; secondly, a test structure has a much smaller gate area in order to avoid process-induced defects; and finally, a large sample size is needed to observe failure times at a very low percentile.

The alternative is to use failure data collected in accelerated life tests to project the reliability under the operating conditions.

Accelerated life testing (ALT) is widely used in the reliability field so that failure information can be collected promptly [45]. Because the failure rate is extremely low under normal working conditions, it is necessary to take actions to force devices to fail within practical time scales. In ALT, devices are subjected to higher than normal level of a stress, e.g., voltage, temperature, humidity, corrosion, radiation, etc., to speed up the degradation process [28]. The acceleration factor of an accelerated stress is defined as

$$AF = \frac{t_{\mathrm{op}}}{t_{\mathrm{ac}}},$$

where t_{ac} is the failure time under the accelerated stress and t_{op} is the corresponding failure time under the normal operating condition.

There are two assumptions to bear in mind when using ALT data to project reliability under normal operating conditions. The first one requires that the failure mechanisms at the two stress levels be identical. This is because the acceleration factor is calculated from some lifetime–stress relationship based on the underlying physical process. The second assumption is that the acceleration factor is linear under an accelerated stress; in other words, the AF at a specific stress level is constant for any possible value of t_{ac}. Figure 16 displays the relationship between the failure distributions under the normal operating condition and an accelerated stress when both assumptions are satisfied.

The two assumptions have significant implications in the case of gate dielectrics, for which $t_{\mathrm{bd}} \sim \mathrm{W}(\alpha, \beta)$. While α is the 63.2^{th} percentile life, β is a measure of the critical defect density, which is unique for each failure mechanism [8, 9]. Assume

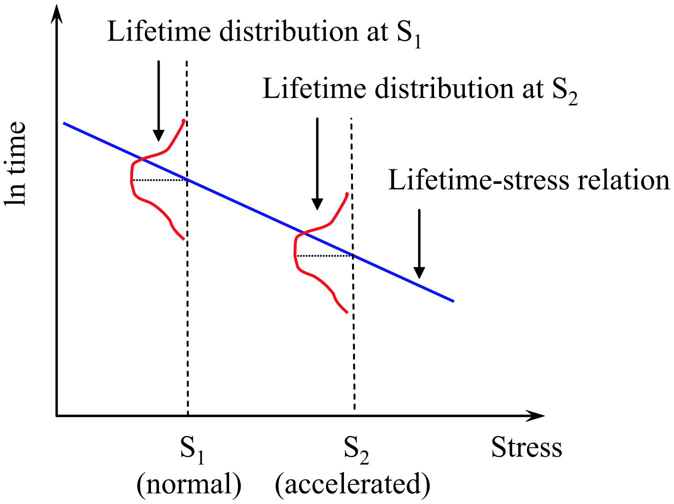


Figure 16: Relationship between the failure distributions under normal operating and accelerated stress conditions

$t_{\mathrm{ac}} \sim \mathrm{W}(\alpha_{\mathrm{ac}}, \beta_{\mathrm{ac}})$ and $t_{\mathrm{op}} \sim \mathrm{W}(\alpha_{\mathrm{op}}, \beta_{\mathrm{op}})$; then by the previous two assumptions,

$$\beta_{\mathrm{op}} = \beta_{\mathrm{ac}}, \quad \frac{\alpha_{\mathrm{op}}}{\alpha_{\mathrm{ac}}} = AF. \tag{4.6}$$

On a Weibull plot, these two distributions are represented as two parallel lines.

IV.4. Physical acceleration models

For gate dielectric thin films, the most significant stress factor is the electric field E across the film. There are several models in the literature that have been used successfully to model the acceleration of electric field for various failure mechanisms. These models are generally written in a deterministic form that says the time-to-breakdown is an exact function of the imposed stress and some material and process dependent constants [59]. However, breakdown is a random event that can not be exactly predicted in advance. Because of the linear acceleration assumption, an acceleration model can be interpreted as an equation that calculates the characteristic life α, or any other percentile life, as a function of the applied stress.

It is generally agreed that the breakdown time of thin oxides follows an exponential law of the electric field, but the specific operative mechanisms and the equations that define them are matters of controversy. The $1/E$ model and the E model have emerged in recent years as the two most popular ways of describing the field dependence of breakdown [48].

The $1/E$ model is derived according to the anode hole injection model. As electrons are injected into the oxide, they undergo impact ionization which generates holes in the process. The slow-moving holes become trapped in the oxide near the cathode, distorting the band diagram and increasing the local electric field. Breakdown occurs when Q_{bd}, the critical density of hole charge, is reached.

$$\alpha = \frac{Q_{\mathrm{bd}}}{j},$$

where j is the tunnelling current depending strongly on E,

$$j = a \cdot \exp(-\frac{G_R}{E}).$$

Then t_{bd} can be written as a function of $1/E$,

$$\alpha = \frac{Q_{\mathrm{bd}}}{a} \exp(\frac{G_R}{E}) = \tau_R \exp(\frac{G_R}{E}), \tag{4.7}$$

where τ_R and G_R are temperature-dependent constants that need to be estimated by other means.

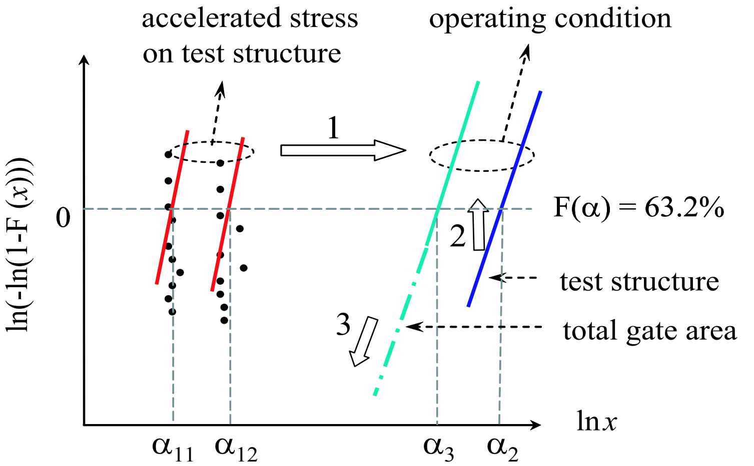
In contrast to the $1/E$ model, the E model suggests

$$\alpha = \tau_L \exp[G_L(E_{\mathrm{bd}} - E)], \tag{4.8}$$

where τ_L and G_L are unknown, temperature-dependent constants. Here E is the applied field and E_{bd} is the field above which breakdown will occur immediately. Unlike the physical basis underlying the $1/E$ model, justification for the E model is largely empirical. The difference between the two models in projecting lifetimes at low voltages is dramatic. Specifically, the $1/E$ model gives optimistic projections while the E model gives pessimistic ones.

IV.5. Procedures of reliability projection

The correctness of reliability projection depends completely on the validity of the extrapolation law, which includes the acceleration model, area scaling, and low percentile extrapolation. Figure 17 illustrates the process of projecting a low percentile lifetime of the total gate oxide in a device at operating condition from ALT data on a test structure [62]. Details of the main procedures are as follows.



1. Stress acceleration 2. Area scaling 3. Percentile extrapolation

Figure 17: Graphic procedures of gate dielectric reliability projection

1. Estimation of the Weibull slope

As aforementioned, the Weibull slope β is a measure of the critical defect density that triggers breakdown for a particular mechanism. It depends on the film's thickness and quality. Because thinner oxides have a larger statistical spread of the average critical defect density that forms the conducting path between the cathode and anode, the Weibull slope decreases with oxide thickness [62]. This has several immediate effects on the accuracy of reliability projection for thin and ultra-thin films. Firstly, area scaling and low percentile extrapolation become much more sensitive to small values of β; secondly, a small error in the estimation of β can raise the error in the lifetime projection to decades [35].

β is difficult to measure to satisfactory accuracy without a very large number of samples [62, 63]. It has been demonstrated that a tight confidence interval of β for a given distribution can be achieved with a sample size of 1000 or more. However, this is not practical in most reliability evaluations. A more accurate method for determining β is to use the relationship given in equation (4.5) [62], but in a different form of

$$\ln \alpha \propto -\frac{1}{\beta} \ln A.$$

If the same stress condition is applied to test structures with different gate sizes, the Weibull failure distribution for each gate size has the same β but a different α. Bu using estimates of the α's and fitting the plot of $\ln \alpha$ versus $\ln A$ with a straight line, then β can be estimated as the reciprocal of the line's slope.

2. Projection of the characteristic life under the normal operating condition

The common way of estimating α_{op}, the characteristic life at the normal operating field, is through ALT. At an accelerated electric field E, breakdown data are fitted with a Weibull distribution, for which α can be estimated by means of least square regression or maximum likelihood estimation. Depending on the acceleration model used, $\ln \alpha$ is linearly regressed against $1/E$ or E. Then $\ln \alpha_{\mathrm{op}}$ can be obtained by extrapolating the fitted line to the low operating field. This step is legitimate under the assumption of identical failure mechanisms at the stress levels involved.

3. Scaling to a total gate area

In the previous two steps, the Weibull slope β, and the characteristic life α_{op1} of a test structure with the gate size A_1 were obtained. Meanwhile, the straight line that represents $\mathrm{W}(\alpha_{\mathrm{op1}}, \beta)$ is put in the Weibull plot. By the relationship given in equation (4.4), shifting this line up the distance of $\ln(A_2/A_1)$ gives the Weibull line for the gate size A_2 of the total gate oxide in the device. Equivalently, the characteristic life α_{op2} corresponding to A_2 can be calculated from equation (4.5),

$$\alpha_{\mathrm{op2}} = \alpha_{\mathrm{op1}}(\frac{A_1}{A_2})^{-\frac{1}{\beta}}.$$

4. Extrapolation to a low percentile life

Because a large number of breakdown events is unacceptable in practice, reliability specifications on gate oxides are usually defined in terms of a very low percentile lifetime. The last step of the projection is to extrapolate the line that represents $\mathrm{W}(\alpha_{\mathrm{op2}}, \beta)$ on the Weibull plot, down to the low percentile of concern.

CHAPTER V

BAYESIAN APPROACH TO RELIABILITY PROJECTION

V.1. General Bayesian approach

In statistics, Bayesian approaches are fundamentally different from classical approaches that people usually take. The classical interpretation of probability is the frequency at which an event occurs; the Bayesian interpretation is that probability represents a person's degree of belief about a certain thing of interest. The essential characteristic of a Bayesian approach is the explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis [18]. This makes Bayesian approaches quite useful with other statistical approaches.

In a classical frequentist approach, a distribution parameter θ is thought to be an unknown, but fixed, quantity. When a random sample X is drawn from the population, knowledge about the value of θ is obtained from the observation. In a Bayesian approach, θ is considered to be a quantity whose variation can be described by a prior distribution, $p(\theta)$. The prior is based on the experimenter's belief and formulated before sample data are seen. A sample is then taken, and the prior distribution is updated according to the Bayes' rule,

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)},$$

where $p(x|\theta)$ is the likelihood function; $p(\theta|x)$ is the posterior distribution, which θ is believed to follow. With $p(\theta|x)$, the Bayesian inference about θ is made in terms of probability statements which are conditioned on the observed value x.

A Bayesian approach is valuable when sample data is limited. On these occasions, information from historical data and/or engineering knowledge can be incorporated into a prior distribution, which is a current educated guess of θ. The posterior distribution actually combines information from the prior and from the sample data. The more sample data that is collected, the more weight it has on the posterior. Even with no data, the Bayesian approach can still provide an estimation that is better than the prior, while classical methods such as maximum likelihood estimation can do nothing in this situation.

V.2. The Bayesian approach with accelerated test data

A Bayesian approach provides a natural way to include uncertainty factors and poor information into reliability modeling. As neither the material science nor the electric properties of high k dielectrics are understood nearly as well as those of SiO_2, the reliability study of high k thin films are built on a foundation of current knowledge about SiO_2. Because the understanding of high k films is very limited at the current time, a Bayesian approach is advantageous over other known methods because it can effectively utilize engineering judgment along with sample data.

For highly-reliable devices, the sample data is likely to be just a few data points even under accelerated stress conditions. Since a maximum likelihood estimator relies heavily on its asymptotic properties, it does not provide a good solution to this kind of problem. However, reliability inference through a Bayesian approach incorporates the prior knowledge about SiO_2 with real test data on high k films and combines them in a way that best represents one's confidence in each.

The common method of lifetime projection requires a large amount of failure data to be collected in ALT at several selected stress levels, so that the maximum likelihood estimate of the Weibull parameters α and β can be found at each level.

Because of the identical failure mechanism assumption of ALT, the electric fields used in CVS tests should be chosen close to the operating conditions in order to reduce extrapolation errors. This results in long stress durations which may be months or more than a year. Even with automatic parallel test equipment, the cost is prohibitively high due to the large amount of resources needed. Besides, the prediction often comes too late to be useful for decision-makings at design stage and during early production.

Reliability projection through accelerated tests is a very difficult problem, for there is no sample data under the normal operating conditions. A Bayesian approach is advantageous in handling this kind of situation because of its less deterministic aspects and the possibility of refinements from previous experiments. A short review of the works dealing with Bayesian methods and accelerated life tests can be found in the literature [14]. In most of these research, Bayesian approaches are applied to estimate the failure rate of an exponential lifetime distribution with ALT data [21, 49]. Only a few papers deal with the flexible and realistic Weibull distribution [11, 14].

In a recent study, Bayesian approach is applied to a Weibull model when a limited amount of failure data from ALT is available [14]. It is assumed that the accelerated stress levels do not change the shape parameter which characterizes the failure mechanism ; and that the scale parameter decreases as an inverse power function of the imposed stress. On the basis of physical and chemical considerations, each model parameter is given an independent truncated uniform prior.

V.3. Bayesian model of lifetime projection

The model proposed for the reliability projection of high k dielectric thin films integrates a statistical lifetime distribution, an empirical acceleration function, and a Bayesian hierarchical framework. The three aspects of the model are explained below in order. Before further discussion, the needed notation is given as follows.

Notation

E_i: the i^{th} level of electric field applied in CVS tests, $i = 1, \cdots, L$

$t_{i,j}$: lifetime of the j^{th} unit under E_i in a CVS test, $j = 1, \cdots, n_i$

$\mathbf{t}_i$: $(t_{i,1}, \cdots, t_{i,n_i})$, $i = 1, \cdots, L$

$\mathbf{T}$: $(t_{1,1}, \cdots, t_{i,n_1}, \cdots, t_{L,1}, \cdots, t_{L,n_L})$

β: shape parameter of the Weibull lifetime distribution

α_i: scale parameter of the Weibull lifetime distribution under E_i

a: intercept of the acceleration function $\ln \alpha_i = a + bE_i + \delta_i$

b: gradient of the acceleration function $\ln \alpha_i = a + bE_i + \delta_i$

a_i: intercept of the acceleration function $\ln \alpha_i = a_i + b_i E_i + \delta_i$, $i = 1, \cdots, L$

b_i: gradient of the acceleration function $\ln \alpha_i = a_i + b_i E_i + \delta_i$, $i = 1, \cdots, L$

δ_i: error term of $\ln \alpha_i = a + bE_i + \delta_i$ or $\ln \alpha_i = a_i + b_i E_i + \delta_i$, $i = 1, \cdots, L$

First of all, the lifetime distribution under a constant stress is assumed to be Weibull, W(α, β), where α and β are to be estimated. It is generally accepted that every distinct failure mechanism corresponds to a unique value of β. But even if the failure mechanism is known and it is identical under different stress conditions, fitting to a Weibull distribution is not always statically valid. In addition, the fitted β depends on the amount and quality of the sample data. Now consider β as a random variable and assign it an appropriate prior. The uncertainty associated with β due to lack of information is represented in this prior distribution

Secondly, assume the E model as the empirical acceleration function. According to the equation (4.8), the model can be written as $\ln \alpha = a + bE$, where a and b are unknown and must be estimated. In the Bayesian context, a and b are treated as random variables instead of fixed, but unknown, constants, and the model takes the form of

$$\ln \alpha = a + bE + \delta,$$

where δ serves as the error term of the model. It is noteworthy that the model parameters are interpreted as random variables in their own right. The stochastic nature of a and b can be interpreted as signifying that the difference in behavior among individual sample capacitors causes the variation in acceleration. This is justified by the fact that test units do not react in the same way to the stress and they do not suffer the same acceleration. By assigning a prior distribution to a and b each, its inherent randomness can be incorporated into the Bayesian model with the uncertainty due to lack of information.

The random variables a, b and δ may or may not depend on the stress level. If the acceleration due to the electric field is almost identical across the stress levels tested, a simplistic model can be assumed in which a and b are independent of the stress level, but δ is not. This is given as

$$\text{Function I}: \quad \ln \alpha_i = a + bE_i + \delta_i. \tag{5.1}$$

If the acceleration changes with the electric field, then a, b and δ should be defined at each stress level. This corresponds to

$$\text{Function II}: \quad \ln \alpha_i = a_i + b_iE_i + \delta_i. \tag{5.2}$$

a_i's and b_i's allow the model to explore regions of very small likelihood and thus tend to result in low posterior probability. However, Function II leads to a large number of parameters as tests are performed at stress levels from E_1 to E_L. The comparison between Function I and Function II in Section VI.4 will show that Function I is

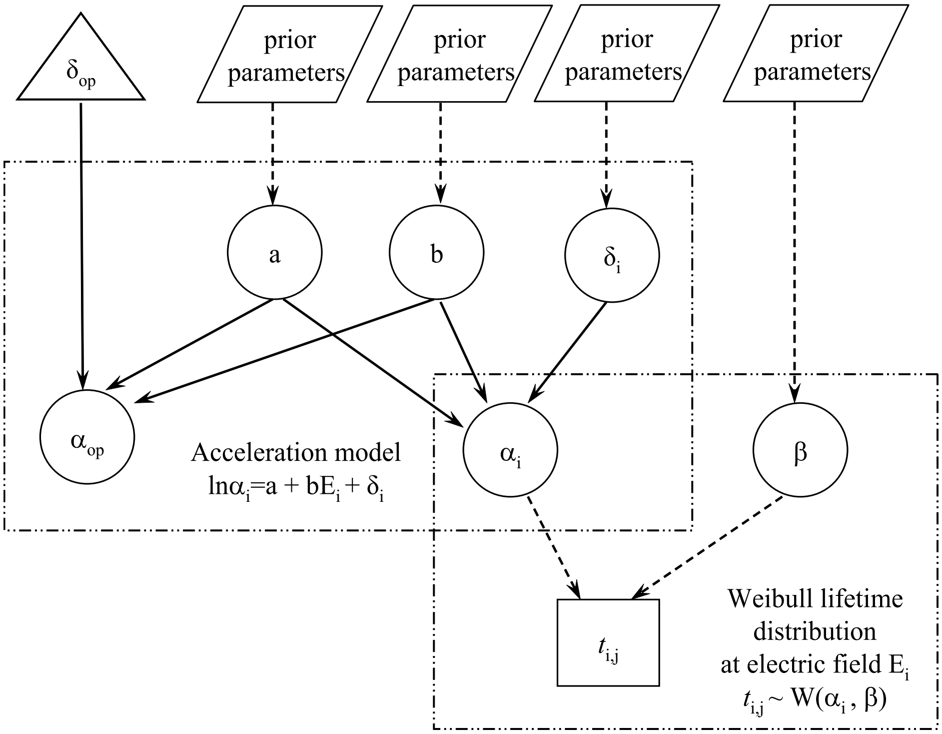


Figure 18: Graphical model of the Bayesian lifetime projection

adequate and convenient. For this reason, the discussion in subsequent sections and chapters assume the use of Function I, except when stated otherwise.

Finally, the hierarchical structure of the Bayesian model is illustrated in Figure 18, in which the circle represents an unknown random variable; the solid square represents observed data; the parallelogram represents fixed parameters of a prior distribution; the dashed arrow represents a probabilistic dependence, while the solid arrow represents a deterministic relationship. In addition, the random variable in the triangle is given a distribution by assumption. The sample data $t_{i,j}$'s are independent conditioning on α_i and β, i.e., $t_{i,j} \sim \mathrm{W}(\alpha_i, \beta)$; α_i depends on a, b and δ_i in the form of $\ln \alpha_i = a + bE_i + \delta_i$. Based on engineering judgment, mutual independence is assumed among the model parameters β, a, b and δ_i's, each of which has its own prior.
]

Let $\pi(\beta)$, $\pi(a)$, $\pi(b)$, and $\pi(\delta_i)$ be the prior densities on β, a, b, and δ_i, respectively. With the previous setting, the marginal density of $\mathbf{t}_i = (t_{i,1}, \cdots, t_{i,n_i})$ is calculated as

$$f(\mathbf{t}_i) = \iiiint f(\mathbf{t}_i|\beta, a, b, \delta_i)\pi(\beta)\pi(a)\pi(b)\pi(\delta_i) \cdot d\beta \cdot da \cdot db \cdot d\delta_i, \tag{5.3}$$

and the marginal density of $\mathbf{T} = \{t_{i,j}, j = 1, \cdots, n_i; i = 1, \cdots, L\}$ is

$$f(\mathbf{T}) = \iiiint \cdots \int f(\mathbf{T}|\beta, a, b, \delta_1, \cdots, \delta_L)\pi(\beta)\pi(a)\pi(b) \cdot d\beta \cdot da \cdot db \cdot \prod_{i=1}^{L} \pi(\delta_i) d\delta_i. \tag{5.4}$$

The posterior densities of β, a , b, and δ_i are given as

$$\pi(\beta|\mathbf{T}) = \frac{\pi(\beta)\iiint \cdots \int f(\mathbf{T}|\beta, a, b, \delta_1, \cdots, \delta_L)\pi(a)\pi(b) \cdot da \cdot db \cdot \prod_{i=1}^{L}\pi(\delta_i)d\delta_i}{f(\mathbf{T})}, \tag{5.5}$$

$$\pi(a|\mathbf{T}) = \frac{\pi(a)\iiint \cdots \int f(\mathbf{T}|\beta, a, b, \delta_1, \cdots, \delta_L)\pi(\beta)\pi(b) \cdot d\beta \cdot db \cdot \prod_{i=1}^{L}\pi(\delta_i)d\delta_i}{f(\mathbf{T})}, \tag{5.6}$$

$$\pi(b|\mathbf{T}) = \frac{\pi(b)\iiint \cdots \int f(\mathbf{T}|\beta, a, b, \delta_1, \cdots, \delta_L)\pi(\beta)\pi(a) \cdot d\beta \cdot da \cdot \prod_{i=1}^{L}\pi(\delta_i)d\delta_i}{f(\mathbf{T})}, \tag{5.7}$$

$$\pi(\delta_i|\mathbf{t}_i) = \frac{\pi(\delta_i) \iiint f(\mathbf{t}_i|\beta, a, b, \delta_i)\pi(\beta)\pi(a)\pi(b) \cdot d\beta \cdot da \cdot db}{f(\mathbf{t}_i)}. \tag{5.8}$$

The reliability under the operating condition, i.e., $\mathrm{W}(\alpha_{\mathrm{op}}, \beta)$, can be known once the estimates of α_{op} and β are obtained. While β's posterior is derived in the model directly, the distribution of α_{op} can be obtained from the posterior distributions of a and b.

The Bayesian approach to lifetime projection is summarized as a flow chart in Figure 19. A few statements should be made for clarity.

- The accelerated stress E_i should be carefully chosen with the consideration that E_i is high enough to induce breakdown in a reasonable time period of testing, but E_i is not overly high so that the risk of failure mechanism changing is as small as possible. In practice, the rule of thumb is that the stress applied in CVS tests should be less than 80% of the breakdown field E_{bd} detected in RVS tests.

- The posterior distribution and inference of the parameters can be obtained with Bayesian computation tools such as Markov chain Monte Carlo simulation.
- All of the information concerning a parameter is contained in its posterior distribution. In order to find an estimator, a loss function needs to be defined in the first place. The Bayes estimator with respect to the squared-error loss function is widely-used, and it equals the posterior mean.
- The Bayesian approach is iterative. The step of updating the prior to get the posterior can be repeated when new data is collected. The most current posterior can be used as the prior of the next iteration.

V.4. Considerations on prior distributions

Generally, likelihood is less sharply peaked for a small sample size or a large number of parameters. In this case, the prior distribution and the hierarchical model are more important [18]. When there are many data points and only a few parameters, a vague or diffuse prior is favored. It gives an acceptable result and takes less effort than specifying prior knowledge in a probabilistic form. In the absence of real prior information, the non-informative Jeffreys' prior can be used [25]. Jeffreys' prior is quite useful for problems with single or just a few parameters to be estimated. It is, however, very complicated if there are many unknown parameters because of the covariance between the parameters. For this reason, Jeffreys' prior is not suitable for the proposed model.

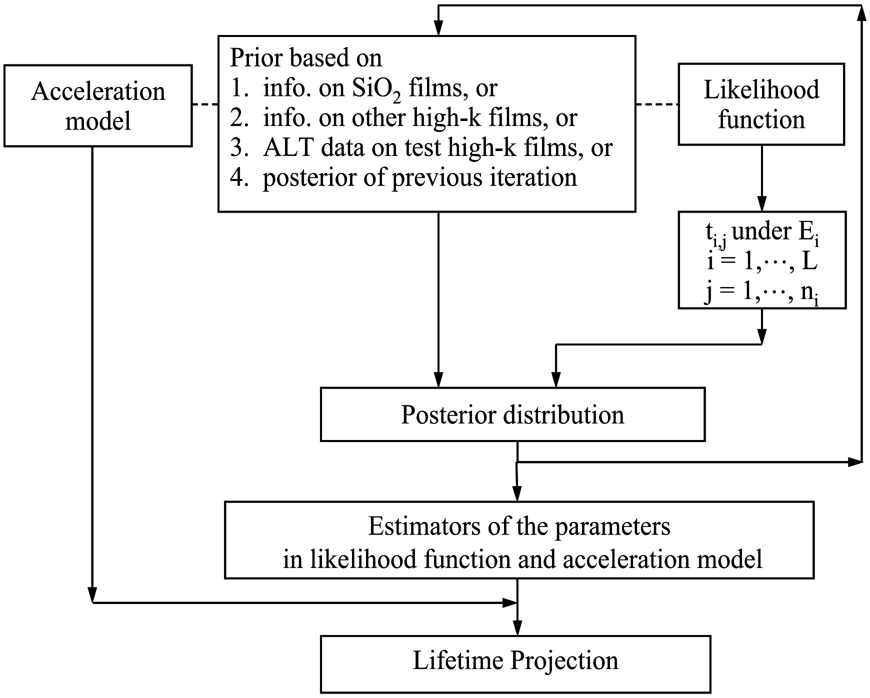


Figure 19: Flow chart of the proposed Bayesian approach

As shown in Figure 19, three sources of information are considered in determining the priors on β, a, b, and δ_i. They are the reports on SiO_2 films in literature, the experimental results on other high k films in recent publications, and the failure data from ALT tests on the high k films under study. From the engineering point of view, dielectric breakdown is sensitive to the structure and fabrication of film stacks. As material, structure and deposition techniques vary among SiO_2 and high k films prepared in different research labs, it is a challenging work to combine information from various sources and utilize it in a systematic way.

When there is less background information on a parameter, a prior with a broad range and flexibility should be suggested to encompass all possible situations and beliefs. To the best of our knowledge, there is not much information concerning the parameters of the projection model available prior to analysis, with the exception of the upper and lower bounds in some cases. In this situation, a vague prior is a good choice. Alternative prior distributions can be selected from truncated uniform, normal, or gamma densities as listed in Table 2.

The parameters of prior distributions are specified on the basis of engineering knowledge and judgement. For instance, β is less than one in the case of a decreasing failure rate and greater than one in the case of an increasing failure rate. There are also β values reported in the literature regarding both SiO_2 and high k films. With this information, a safe range for β is set as $(0.1, 1.9)$. For the acceleration function to be physically meaningful, it is required that the intercept a be positive and the gradient b be negative. Engineering knowledge also indicates that a is within $(1, 50)$ and b within $(-8, -0.2)$, with no preferred values in each range. The error term δ_i is given a normal prior centered at zero with variance in a wide range. In addition, empirical priors can be constructed for future studies if there is a certain amount of historical data available.

V.5. Sample data of breakdown

To collect time-to-breakdown data, CVS tests are performed on the 20W Hf-doped TaO_x film listed in Table 1. The accelerated stresses, ranging from 5.5 to 8.1MV/cm, are less than 80% of the breakdown field E_{bd}, which is measured as above 10MV/cm in the RVS tests. In the CVS tests, the jump of leakage current at the moment of breakdown can be clearly identified without monitoring relaxation current or leakage current at a low field.

All of the test capacitors are randomly allocated to a stress level. Each is tested individually and breaks down independently. For the capacitors that are stressed under E_i, their failure times are independently and identically distributed as

Table 2: Variety of prior distributions being considered

	Case 1	Case 2	Case 3	Case 4
$\pi(\beta)$	U(0.1,1.9)[†]	U(0.1,1.9)	U(0.1,1.9)	U(0.1,1.9)
$\pi(a)$	U(1,50)	U(1,50)	U(1,50)	U(1,50)
$\pi(b)$	U($-8, -0.2$)	U($-8, -0.2$)	U($-8, -0.2$)	U($-8, -0.2$)
$\pi(\delta_i)$	N(0,1/64)[‡]	N(0,1/100)	N(0,1/400)	N(0,1/900)

	Case I	Case II	Case III	Case IV
$\pi(\beta)$	U(0.4,0.9)	U(0.1,1.9)	U(0.1,1.9)	U(0.1,1.9)
$\pi(a)$	U(10,40)	U(1,50)	U(1,50)	G(10,0.5)[♯]
$\pi(b)$	U($-5, -0.5$)	U($-8, -0.2$)	N($-3, 0.4$)	$\pi(-b) \sim$G(3,1)
$\pi(\delta_{55})$[*]	N(0,1/300)	N(0,1/300)	N(0,1/300)	N(0,1/300)
$\pi(\delta_{60})$	N(0,1/300)	N(0,1/300)	N(0,1/300)	N(0,1/300)
$\pi(\delta_{65})$	N(0,1/300)	N(0,1/300)	N(0,1/300)	N(0,1/300)
$\pi(\delta_{71})$	N(0,1/900)	N(0,1/900)	N(0,1/900)	N(0,1/900)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\pi(\delta_{81})$	N(0,1/900)	N(0,1/900)	N(0,1/900)	N(0,1/900)

[*] As an example, δ_{55} denotes δ_i at $E = 5.5$MV/cm.

[†] U(u, v) is a truncated uniform distribution, $f(x; u, v) = \frac{1}{v-u}, u \leq x \leq v$.

[‡] N(μ, σ^2) is a normal distribution, $f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$.

[♯] G(ζ, η) is a gamma distribution, $f(x; \zeta, \eta) = \frac{\eta^\zeta x^{\zeta-1} e^{-\eta x}}{\Gamma(\zeta)}, 0 \leq x < \infty$.

W(α_i, β). The ordered breakdown times at each of the nine electric fields are given in Table 3 and Table 4.

Table 3: Part one of ordered breakdown time in seconds

Order	Electric field E (MV/cm)								
	8.1	7.9	7.7	7.5	7.3	7.1	6.5	6.0	5.5
1	1	1	9	39	10	28	1659	5657	3482
2	5	2	18	40	21	88	3312	30481	34042
3	10	9	20	77	46	99	13111	137907	56526
4	10	12	25	247	55	107	21392	152911	57630
5	10	35	29	253	58	211	–	–	–
6	11	46	66	299	130	213			
7	14	72	124	311	284	248			
8	20	74	127	633	353	301			
9	28	82	175	666	361	311			
10	68	107	221	830	395	593			
11	78	142	249	950	426	673			
12	92	153	341	1060	574	702			
13	93	193	362	1383	739	741			
14	116	251	552	1416	990	911			
15	125	290	630	1742	1004	949			

Table 4: Part two of ordered breakdown time in seconds

Order	Electric field E (MV/cm)								
	8.1	7.9	7.7	7.5	7.3	7.1	6.5	6.0	5.5
16	128	348	760	1843	1321	1040			
17	141	399	782	1879	1336	1439			
18	167	511	794	1905	1631	1971			
19	219	556	906	2096	2040	2069			
20	311	1104	932	2337	2043	2253			
21	407	1509	968	2532	2060	2501			
22	502	1535	1378	2648	2154	3547			
23	506	1756	1386	3020	2181	4452			
24	682	2376	1664	3434	2398	4580			
25	709	2843	1728	3947	2458	4882			
26	772	3140	2229	4373	4210	5657			
27	1209	3514	2249	4729	4702	5737			
28	1282	3616	2338	5215	7461	6323			
29	1485	3882	4058	5614	8011	7565			
30	1638	4583	4986	6753	8424	8209			
31	2154	–	6312	9703	8681	10000			
32	2443		6400	9898	10070	11650			
33	–		6847	10130	13350	15250			
34			8474	11090	22460	21620			
35			–	–	–	25910			

CHAPTER VI

POSTERIOR SIMULATION AND INFERENCE

When a Bayesian approach is applied to high-dimensional problems, multiple levels of integration are necessary to obtain the normalizing constant and marginal densities of the model's parameters. For complicated models, these marginal densities are often analytically intractable, and sometimes even a numerical integration cannot be directly obtained. In these cases, Markov chain Monte Carlo simulation (MCMC) is the easiest way to get reliable results without evaluating integrals [17]. The idea is to create a Markov process whose stationary distribution is the specified target distribution, then run the simulation long enough so that the distribution of simulated draws closely approximates the stationary distribution [18].

A MCMC algorithm that is particularly useful in high-dimensional problems is the alternating conditional sampling called Gibbs sampler [17]. Each iteration of the Gibbs sampler cycles through the unknown parameters, drawing a sample of one parameter conditional on the latest value of all the others. When the number of iterations is large enough, the sample draws on one parameter can be regarded as simulated observations from its marginal distribution. Hence, the marginal density is reconstructed by averaging over the conditional density of the sample draws on this parameter. This marginal distribution is the posterior of the parameter and can easily be used to make inferences on the functions of the model parameters.

The Gibbs sampler reduces the problem of dealing simultaneously with a large number of intricately related parameters to a simple problem of sampling one at a time from the full conditional distribution. It is straightforward to specify the distribution functions required for Gibbs sampling, especially if the model is constructed using a sequence of conditional probability functions as in the proposed hierarchical model. WinBUGS, a Windows version of the Bayesian inference using Gibbs sampling, is a specialized software for implementing MCMC and Gibbs sampling [20, 53]. As a convenient tool for Bayesian analysis, it is used in this study for posterior inference.

VI.1. Full conditional densities

An unknown random variable is called a stochastic node in WinBUGS. Its full conditional distribution is the probability density function conditioning on the current values of all other stochastic nodes in the model. For a specific stochastic node, the full conditional density is proportional to the product of its prior density with the likelihood of all the other nodes that have conditional dependence on it. In Gibbs sampling, simulated draws are sampled from full conditional distributions, but WinBUGS does not require explicit evaluation of the integration constant of a full conditional density function. The full conditional densities of the parameters in the proposed model are derived as follows for reference.

Under the electric field E_i, $t_{i,j}$'s follow $\mathrm{W}(\alpha_i, \beta)$ with the density function of

$$f(t_{i,j}|\alpha, \beta) = \frac{\beta}{\alpha}(\frac{t_{i,j}}{\alpha})^{\beta-1} \exp[-(\frac{t_{i,j}}{\alpha})^{\beta}].$$

According to the acceleration function I: $\ln \alpha_i = a + bE_i + \delta_i$, α_i can be written as

$$\alpha_i = e^{a+bE_i+\delta_i}.$$

Then the sampling density function of $t_{i,j}$ in terms of β, a, b and δ_i has the form of

$$f(t_{i,j}|\beta, a, b, \delta_i) = \beta e^{-\beta(a+bE_i+\delta_i)} t_{i,j}^{\beta-1} \exp(-e^{-\beta(a+bE_i+\delta_i)} t_{i,j}^{\beta}). \tag{6.1}$$

Write $\mathbf{T} = \{t_{i,j}, j = 1, \cdots, n_i; i = 1, \cdots, L\}$ and $\Delta = (\delta_1, \cdots, \delta_L)$. The joint density function of the sample data $\mathbf{T}$ and the model parameters, β, a, b and Δ, can be written as

$$f(\mathbf{T}, \beta, a, b, \Delta) = \pi(\beta)\pi(a)\pi(b) \prod_{i=1}^{L} [f(\delta_i) \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i)]. \tag{6.2}$$

The full conditional density of each parameter is derived as follows.

$$f(\beta|\mathbf{T}, a, b, \Delta) \propto \pi(\beta) \cdot \prod_{i=1}^{L} \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i). \tag{6.3}$$

$$f(a|\mathbf{T}, \beta, b, \Delta) \propto \pi(a) \cdot \prod_{i=1}^{L} \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i). \tag{6.4}$$

$$f(b|\mathbf{T}, \beta, a, \Delta) \propto \pi(b) \cdot \prod_{i=1}^{L} \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i). \tag{6.5}$$

$$f(\delta_i|\mathbf{T}, \beta, a, b, \Delta_{-i}) \propto \pi(\delta_i) \cdot \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i), i = 1, \cdots, L, \tag{6.6}$$

where $\Delta_{-i} = (\delta_1, \cdots, \delta_{i-1}, \delta_{i-1}, \cdots, \delta_L)$. The product term in the right hand side of (6.6) is

$$\prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i) = \beta^{n_i} (\prod_{j=1}^{n_i} t_{i,j}^{\beta-1}) e^{-\beta n_i (a+bE_i+\delta_i)} \exp(-e^{-\beta(a+bE_i+\delta_i)} \sum_{j=1}^{n_i} t_{i,j}^{\beta});$$

and the product term in the right hand side of (6.3), (6.4), and (6.5) is

$$\begin{aligned} \prod_{i=1}^{L} \prod_{j=1}^{n_i} f(t_{i,j}|\beta, a, b, \delta_i) &= \prod_{i=1}^{L} \prod_{j=1}^{n_i} t_{i,j}^{\beta-1} \cdot \prod_{i=1}^{L} \beta^{n_i} e^{-\beta n_i (a+bE_i+\delta_i)} \\ &\quad \cdot \exp[-\sum_{i=1}^{L} (e^{-\beta(a+bE_i+\delta_i)} \sum_{j=1}^{n_i} t_{i,j}^{\beta})]. \end{aligned}$$

Next, substitute $\pi(\beta)$, $\pi(a)$, $\pi(b)$ and $\pi(\delta_i)$ with the density functions given in Table 2 to see the actual form of the full conditionals. For simplicity, the same parameter notation is used for one type of distribution functions just to show the form of its density function. For instance, a uniform is denoted as $\mathrm{U}(u,v)$, a normal as $\mathrm{N}(\mu,\sigma^2)$, and a gamma as $\mathrm{G}(\zeta,\eta)$. For the same reason, there is no need to distinguish the notation of known constants, s_1, s_2, s_3, and w_i's, between different formulas.

<u>Case I and II</u>: $\beta \sim \mathrm{U}(u,v), a \sim \mathrm{U}(u,v), b \sim \mathrm{U}(u,v), \delta_i \sim \mathrm{N}(\mu,\sigma^2)$

$$f(\beta|\mathbf{T},a,b,\Delta) \propto \beta^{s_1} s_2^{\beta} \exp[s_3\beta - \sum_{i=1}^{L}(e^{w_i\beta}\sum_{j=1}^{n_i} t_{i,j}^{\beta})]. \tag{6.7}$$

$$f(a|\mathbf{T},\beta,b,\Delta) \propto \exp(s_1 a + s_2 e^{-\beta a}). \tag{6.8}$$

$$f(b|\mathbf{T},\beta,a,\Delta) \propto \exp(s_1 b + s_2 e^{-\beta E_i b}). \tag{6.9}$$

$$f(\delta_i|\mathbf{T},\beta,a,b,\Delta_{-i}) \propto \exp(s_1\delta_i^2 + s_2\delta_i + s_3 e^{-\beta\delta_i}). \tag{6.10}$$

The right hand side of each of (6.7), (6.8), (6.9) and (6.10) is a log-concave function on β, a, b and δ_i, respectively.

<u>Case III</u>: $\beta \sim \mathrm{U}(u,v), a \sim \mathrm{U}(u,v), b \sim \mathrm{N}(\mu,\sigma^2), \delta_i \sim \mathrm{N}(\mu,\sigma^2)$

$f(\beta|\mathbf{T},a,b,\Delta)$, $f(a|\mathbf{T},\beta,b,\Delta)$ and $f(\delta_i|\mathbf{T},\beta,a,b,\Delta_{-i})$ are in the same form as (6.7), (6.8) and (6.10), respectively.

$$f(b|\mathbf{T},\beta,a,\Delta) \propto \exp(s_1 b + s_2 b^2 + s_3 e^{-\beta E_i b}). \tag{6.11}$$

The right hand side of each of (6.11) is a log-concave function on b.

<u>Case IV</u>: $\beta \sim \mathrm{U}(u,v), a \sim \mathrm{G}(\zeta,\eta), b \sim \mathrm{G}(\zeta,\eta), \delta_i \sim \mathrm{N}(\mu,\sigma^2)$

$f(\beta|\mathbf{T},a,b,\Delta)$ and $f(\delta_i|\mathbf{T},\beta,a,b,\Delta_{-i})$ have the same form as (6.7) and (6.10), respectively.

$$f(a|\mathbf{T},\beta,b,\Delta) \propto a^{\zeta-1}\exp(s_1 a + s_2 e^{-\beta a}). \tag{6.12}$$

$$f(b|\mathbf{T},\beta,a,\Delta) \propto b^{\zeta-1}\exp(s_1 b + s_2 e^{-\beta E_i b}). \tag{6.13}$$

The right hand side of (6.12) and (6.13) are log-concave if $\zeta > 1$.

VI.2. Simulation with WinBUGS

The code generator in WinBUGS contains a small expert system for deciding the best sampling method for a stochastic node according to its full conditional density [20]. Sampling methods are applied in the hierarchies outlined in Table 5. A sampling method is used if no previous method in the hierarchy is appropriate. The sampling method that WinBUGS selects for the stochastic nodes of the projection model are listed in Table 6.

Table 5: Hierarchy of sampling methods in WinBUGS

	Continuous target distribution	Sampling method
1	Conjugate	Direct sampling using standard algorithms
2	Log-concave	Derivative-free adaptive rejection sampling
3	Restricted range	Slice sampling
4	Unrestricted range	Current point Metropolis algorithm

Table 6: Sampling methods used for β, a, b and δ_i in WinBUGS

Node	Sampling Method
β	Slice sampling
a	Derivative-free adaptive rejection sampling
b	Derivative-free adaptive rejection sampling
δ_i	Rejection sampling

The Metropolis algorithm in WinBUGS is based on a univariate normal proposal distribution centered at the current value. The standard deviation of the normal proposal is tuned over the first 4000 iterations in order to get an acceptance rate between 20% and 40%. All summary statistics for the model will ignore information from this adapting phase. In WinBUGS, slice sampling is used for non log-concave densities on a restricted range. This method can avoid the need to sample from non-standard distributions [44]. It has an adaptive phase of 500 iterations which will be discarded from all summary statistics.

To prevent posterior dependence on the starting point of a simulation, it is important to run several chains with over-dispersed starting points in one MCMC simulation. The simulation converges to the target distribution when traces of all the chains appear to be mixing together. A quantitative way of monitoring the convergence is based on an analysis of variance. A reasonable convergence is diagnosed when the variance between the different chains is no longer larger than the variance within each individual chain. Based on this, the Gelman-Rubin convergence statistic, R, is defined as the ratio of the width of the central 80% interval of the pooled chains to the average width of the 80% intervals of the individual chains [53]. When a WinBUGS simulation converges, R should be at, or close to, one and both the polled and average widths should appear to be stable. Simulated draws after convergence from all parallel chains are pooled together for posterior analysis and inference, whereas those before convergence are discarded.

The accuracy of a posterior estimate is calculated in terms of a Monte Carlo standard error (MC error) of the mean [53]. The MC error gives an estimate of the difference between the mean of the sampled values and the true posterior mean of the stochastic node. As a rule of thumb, the simulation should run until the MC error for each node is less than 5% of the sample standard deviation (Std.D.).

For the projection model, four chains run simultaneously in one simulation. Each chain continues for 60,000 iterations, and the first 10,000 iterations are discarded before posterior analysis. Convergence of the simulation is reached when the Gelman-Rubin statistic R stays at one after 10,000 iterations. Because of the large number of iterations, the MC errors are much less than 5% of the sample standard deviation.

VI.3. Sensitivity analysis

Posterior inference depends on model assumptions such as sampling distribution and priors. If the choice of the prior has little effect upon the posterior of a parameter, it reflects the fact that the likelihood carries strong information concerning the parameter. However, if there is relatively little information in the likelihood, an unrealistic prior on a parameter results in not only a large change in its posterior, but also changes in other parameter's posterior. In general, models which contain parameters that are not supported by sample data tend to have a high degree of posterior sensitivity; thus the degree of sensitivity also provides evidence in favor of one model over another.

For the projection model, the posterior sensitivity is examined under the priors given in Table 2. The priors of Case 1 to Case 4 are used to examine whether the prior variation of δ_i affects the posteriors of the model parameters; the priors of Case I to Case IV are intended to reveal the influence that the priors of β, a and b have on the posteriors. The posterior statistics and kernel densities of the model parameters are compared using the simulation results of WinBUGS.

Some important observations merge from monitoring the simulation process. The Markov chain of β converges much earlier than that of any other node, and β's posterior is very stable over a large variety of priors. This implies that, although the data are from different electric fields, they indicate similar Weibull shape parameters and hence an identical failure mechanism. Because the sample data are consistent in this respect, a vague prior on β, such as a truncated uniform, can yield a satisfying posterior. However, the Markov chain for either a or b does not reach a good convergence in the first 10,000 iterations. This is due to the high uncertainty as well as the large physical variations in the parameter that result from the information contained in the sample data.

VI.3.1. Comparison among the priors of Cases 1–4

The priors on a, b and β are kept the same from Case 1 to Case 4, but the variance of δ_i decreases following the order of 1/64, 1/100, 1/400 to 1/900. Posterior statistics and kernel densities of δ_i's, β, a, and b can be found in Table 7, Table 8, Figure 20, and Figure 21. Due to space limitations, the posterior statistics of five δ_i's are given in Table 7; the kernel density of δ_{55} in Figure 20 is used as an example for all δ_i's.

Table 7 and Figure 20 show that, in each of the four cases, the posterior of δ_i is very similar to its prior in terms of mean, variance, and the shape of the density function. This tells us that either the prior of δ_i outweighs the likelihood in its posterior or, more likely, the sample data at E_i do not contain much information concerning δ_i.

For each β, a, and b, the difference in the posterior mean or the posterior variance is negligible among the four prior cases. In addition, the posterior kernel density can be fitted with a normal density quite well. All of these results lead to the conclusion that the parameters of the projection model are insensitive to a change in the variation of the error terms.

Table 7: Posterior statistics of δ_i under the priors of Case 1 to Case 4

Node[†]	Mean (10^{-2})	Std.D. (10^{-1})	MC error (10^{-4})	2.5% (10^{-1})	Median (10^{-2})	97.5% (10^{-1})
$\delta_{55}[1]$	-1.989	1.242	3.339	-2.636	-1.991	2.228
$\delta_{55}[2]$	-1.296	0.995	2.498	-2.077	-1.296	1.824
$\delta_{55}[3]$	-0.331	0.500	1.136	-1.012	-0.322	0.950
$\delta_{55}[4]$	-0.144	0.333	0.737	-0.666	-0.150	0.638
$\delta_{60}[1]$	2.367	1.234	4.823	-2.188	2.378	2.652
$\delta_{60}[2]$	1.485	0.989	3.416	-1.792	1.513	2.079
$\delta_{60}[3]$	0.387	0.499	1.295	-0.940	0.397	1.016
$\delta_{60}[4]$	0.168	0.332	0.788	-0.633	0.160	0.671
$\delta_{71}[1]$	-3.176	1.161	5.431	-2.588	-3.180	1.956
$\delta_{71}[2]$	-2.272	0.957	4.100	-2.107	-2.270	1.648
$\delta_{71}[3]$	-0.587	0.495	1.414	-1.028	-0.593	0.911
$\delta_{71}[4]$	-0.283	0.332	0.812	-0.679	-0.275	0.626
$\delta_{75}[1]$	7.708	1.133	2.731	-1.441	7.685	2.999
$\delta_{75}[2]$	5.341	0.932	2.282	-1.291	5.329	2.363
$\delta_{75}[3]$	1.502	0.491	1.140	-0.805	1.468	1.113
$\delta_{75}[4]$	0.676	0.331	0.743	-0.583	6.693	7.183
$\delta_{81}[1]$	-6.208	1.181	6.588	-2.930	-6.227	1.698
$\delta_{81}[2]$	-4.145	0.9676	4.909	-2.311	-4.154	1.478
$\delta_{81}[3]$	-1.153	0.4939	1.494	-1.079	-1.174	0.855
$\delta_{81}[4]$	-0.507	0.332	0.873	-0.701	-0.509	0.601

[†] As an example, δ_{55} denotes δ_i at $E = 5.5\text{MV/cm}$.

[1]-[4] correspond to the priors of Case 1 to Case 4, respectively.

Table 8: Posterior statistics of a, b and β under the priors of Case 1 to Case 4

Node†	Mean	Std.D.	MC error	2.5%	Median	97.5%
$a[1]$	22.64	1.707	4.982×10^{-2}	19.39	22.57	26.12
$a[2]$	22.75	1.767	5.239×10^{-2}	19.43	22.71	26.37
$a[3]$	22.63	1.641	4.752×10^{-2}	19.53	22.60	25.89
$a[4]$	22.72	1.676	4.924×10^{-2}	19.53	22.68	26.12
$b[1]$	-2.022	0.228	6.644×10^{-3}	-2.485	-2.013	-1.588
$b[2]$	-2.037	0.236	6.980×10^{-3}	-2.519	-2.032	-1.594
$b[3]$	-2.021	0.219	6.330×10^{-3}	-2.455	-2.018	-1.608
$b[4]$	-2.032	0.223	6.555×10^{-3}	-2.484	-2.027	-1.608
$\beta[1]$	0.668	3.679×10^{-2}	1.156×10^{-4}	0.598	0.667	0.742
$\beta[2]$	0.668	3.681×10^{-2}	1.138×10^{-4}	0.597	0.667	0.741
$\beta[3]$	0.667	3.666×10^{-2}	1.131×10^{-4}	0.597	0.667	0.741
$\beta[4]$	0.667	3.662×10^{-2}	1.124×10^{-4}	0.597	0.666	0.740

†[1]-[4] correspond to the priors of Case 1 to Case 4, respectively.

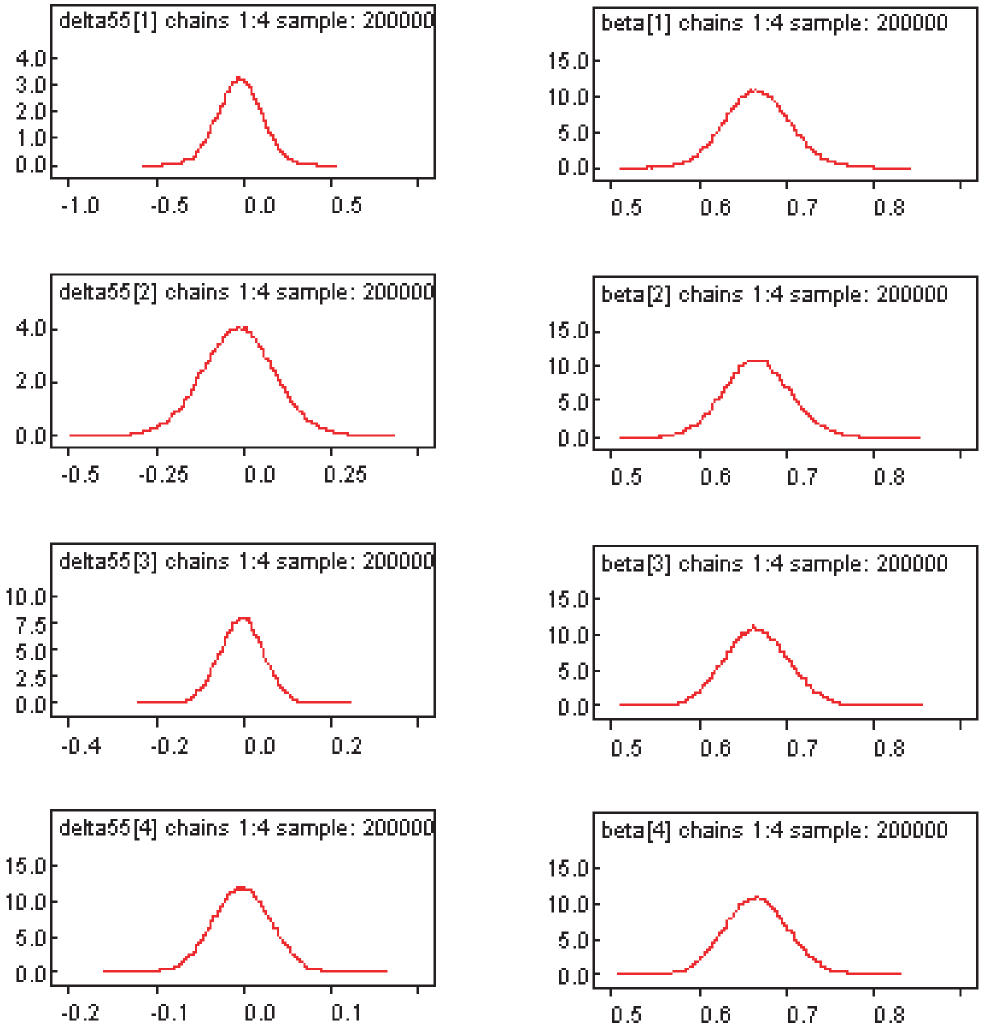


Figure 20: Posterior kernel densities of δ_{55} and β under the priors of Case 1 to Case 4, which are denoted as [1] to [4], respectively.

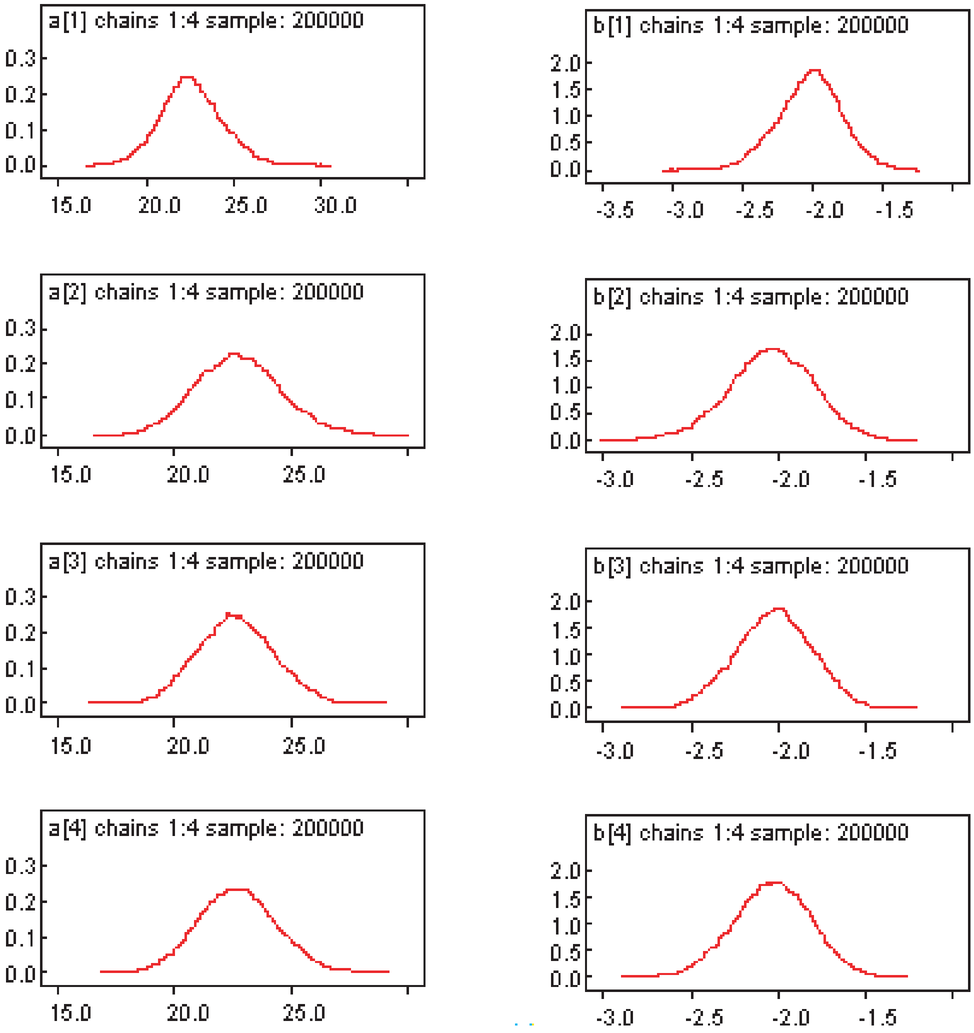


Figure 21: Posterior kernel densities of a and b under the priors of Case 1 to Case 4, which are denoted as [1] to [4], respectively.

VI.3.2. Comparison among the priors of Cases I–IV

Among Case I to Case IV, the priors on β, a, and b vary while the priors on δ_i's are kept same. The difference in the priors of β, a, and b between Case I and Case II is that the truncated intervals are tighter in Case I. From Case II, the prior on b is replaced with a normal density in Case III, while the prior on both a and b is replaced with a gamma density in Case IV. Since the failure time at high stress levels are less spread out, the priors on δ_{71} to δ_{81} are assumed to be normal densities with a smaller variance than those of δ_{50} to δ_{65}. The posterior statistics and kernel densities are given in Table 9, Table 10, Figure 22, and Figure 23.

Irrespective of the prior case, the posterior of each model parameter has a normal-like kernel density, which is similar from Case I to Case IV. The posteriors of a or b from the small-variance normal prior and the right-skewed gamma prior are almost the same as that from the truncated uniform prior. The fact that the model parameters are insensitive to small changes in the priors of a and b has two indications. On the one hand, the sample data put much more weight on the posteriors than does the prior; on the other hand, the prior is sufficiently diffuse so that the effect caused by a small change in it is negligible on the posterior.

VI.4. Comparison between two acceleration functions

In order to see whether the acceleration depends on the stress level of the electric field, a MCMC simulation is also run under Function II: $\ln \alpha_i = a_i + b_i E_i + \delta_i$. The posterior statistics of a_i's and b_i's are compared with those of a and b from Function I under the prior Case I. $\pi(\delta_i)$ is exactly the same for Function I and Function II, whereas both $\pi(a_i)$ and $\pi(b_i)$ for Function II are given the same density as $\pi(a)$ and $\pi(b)$, respectively.

Table 9: Posterior statistics of a, b and β under the priors of Case I to Case IV

Node†	Mean	Std.D.	MC error	2.5%	Median	97.5%
$a[1]$	22.73	1.707	5.056×10^{-2}	19.57	22.67	26.12
$a[2]$	22.65	1.687	4.932×10^{-2}	19.50	22.62	26.08
$a[3]$	22.82	1.661	4.882×10^{-2}	19.71	22.74	26.22
$a[4]$	22.36	1.530	4.360×10^{-2}	19.38	22.32	25.48
$b[1]$	−2.034	0.227	6.730×10^{-3}	−2.485	−2.026	−1.613
$b[2]$	−2.024	0.225	6.568×10^{-3}	−2.480	−2.020	−1.603
$b[3]$	−2.046	0.221	6.499×10^{-3}	−2.499	−2.035	−1.631
$b[4]$	−1.985	0.204	5.809×10^{-3}	−2.401	−1.980	−1.588
$\beta[1]$	0.667	3.669×10^{-2}	1.166×10^{-4}	0.597	0.666	0.741
$\beta[2]$	0.667	3.667×10^{-2}	1.134×10^{-4}	0.597	0.666	0.740
$\beta[3]$	0.667	3.662×10^{-2}	1.172×10^{-4}	0.597	0.667	0.740
$\beta[4]$	0.667	3.667×10^{-2}	1.065×10^{-4}	0.597	0.667	0.741

†[1]-[4] correspond to the priors of Case I to Case IV, respectively.

Table 10: Posterior statistics of δ under the priors of Case I to Case IV

Node†	Mean (10^{-3})	Std.D. (10^{-2})	MC error (10^{-4})	2.5% (10^{-1})	Median (10^{-3})	97.5% (10^{-1})
$\delta_{55}[1]$	-4.362	5.766	1.309	-1.178	-4.500	1.089
$\delta_{55}[2]$	-4.593	5.775	1.317	-1.180	-4.661	1.084
$\delta_{55}[3]$	-4.296	5.776	1.339	-1.173	-4.499	1.094
$\delta_{55}[4]$	-4.230	5.767	1.367	-1.173	-3.971	1.089
$\delta_{60}[1]$	5.257	5.769	1.624	-1.076	5.326	1.182
$\delta_{60}[2]$	5.067	5.751	1.532	-1.079	5.026	1.176
$\delta_{60}[3]$	4.732	5.756	1.588	-1.078	4.756	1.178
$\delta_{60}[4]$	5.651	5.738	1.489	-1.069	5.598	1.182
$\delta_{71}[1]$	-2.983	3.313	0.819	-0.680	-2.897	0.618
$\delta_{71}[2]$	-2.911	3.322	0.847	-0.680	-2.908	0.622
$\delta_{71}[3]$	-3.079	3.308	0.866	-0.678	-3.125	0.617
$\delta_{71}[4]$	-2.644	3.321	0.825	-0.680	-2.611	0.623
$\delta_{75}[1]$	6.830	3.315	0.729	-0.584	6.875	0.717
$\delta_{75}[2]$	6.793	3.304	0.747	-0.581	6.899	0.717
$\delta_{75}[3]$	6.877	3.299	0.759	-0.576	6.754	0.718
$\delta_{75}[4]$	6.775	3.310	0.753	-0.580	6.706	0.718
$\delta_{81}[1]$	-5.167	3.330	0.877	-0.702	-5.210	0.603
$\delta_{81}[2]$	-5.078	3.326	0.879	-0.703	-4.999	0.602
$\delta_{81}[3]$	-5.078	3.320	0.866	-0.701	-5.011	0.600
$\delta_{81}[4]$	-5.499	3.326	0.869	-0.707	-5.385	0.5962

†As an example, δ_{55} denotes δ_i at $E = 5.5$MV/cm.

[1]-[4] correspond to the priors of Case I to Case IV, respectively.

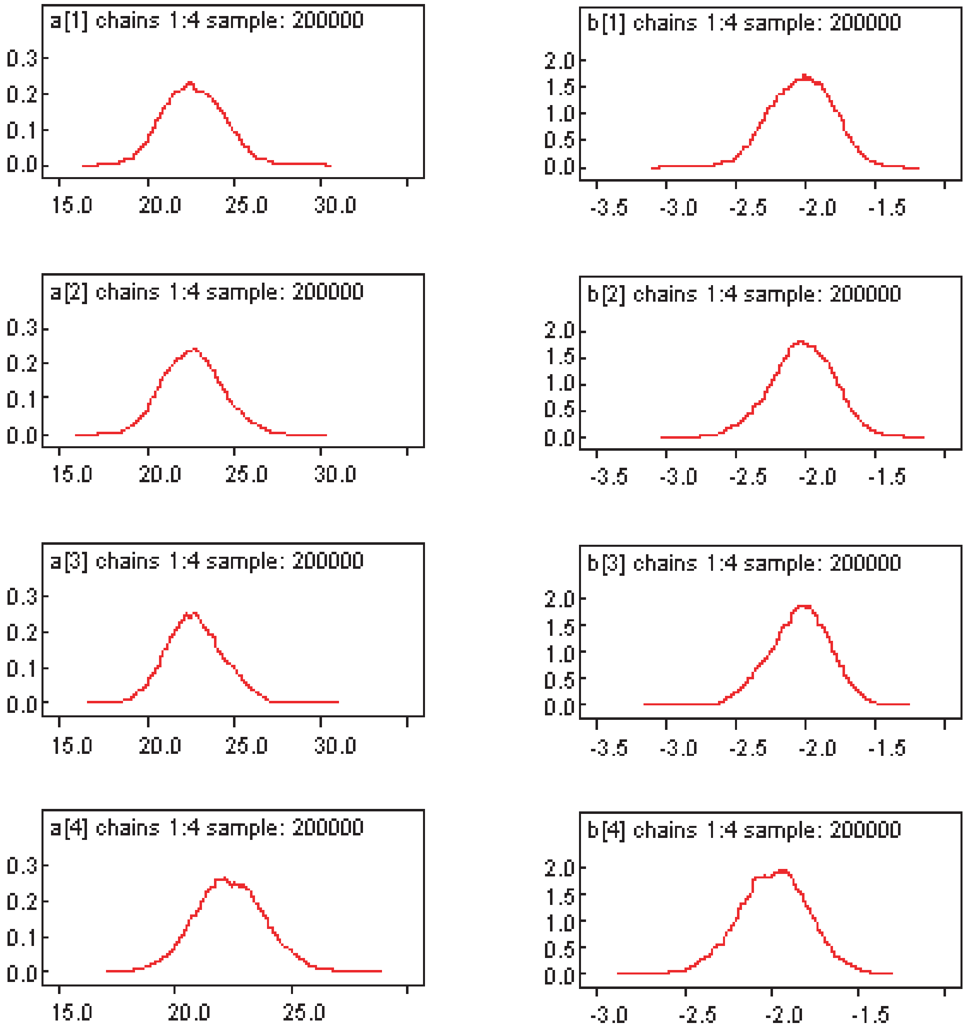


Figure 22: Posterior kernel densities of a and b under the priors of Case I to Case IV, which are denoted as [1] to [4], respectively.

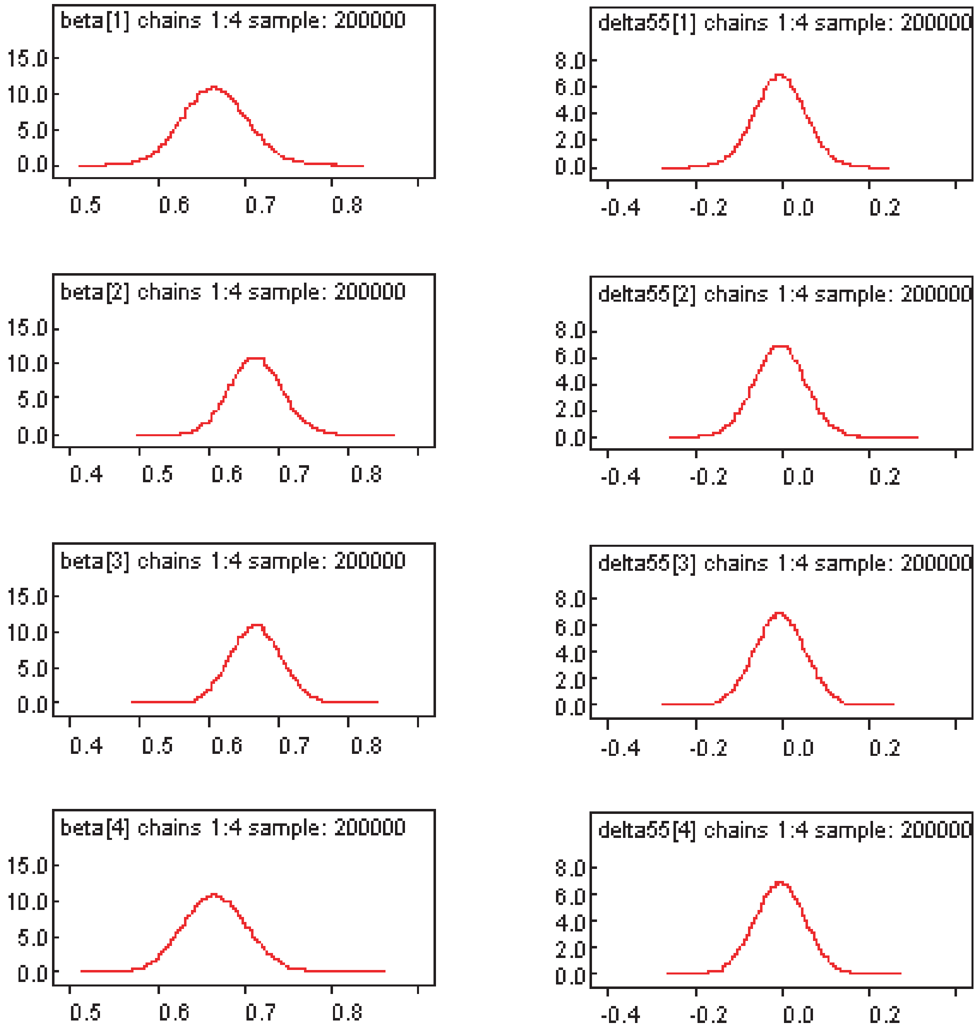


Figure 23: Posterior kernel densities of β and δ_{55} under the priors of Case I to Case IV, which are denoted as [1] to [4], respectively.

Notice that the posterior kernel density of either a_i or b_i is similar to its prior of a truncated uniform. As examples, the kernel densities of some a_i's and b_i's are given in Figure 24. These kernel densities are quite different from the normal-like densities of a and b shown in Figure 22. The reason is that the sample size used for estimating a_i and b_i is reduced greatly from that used for a and b. Consequently, the prior of a_i or b_i has relatively more weight on the posterior, and the posterior has a large variance compared with that of the a or b derived under Function I.

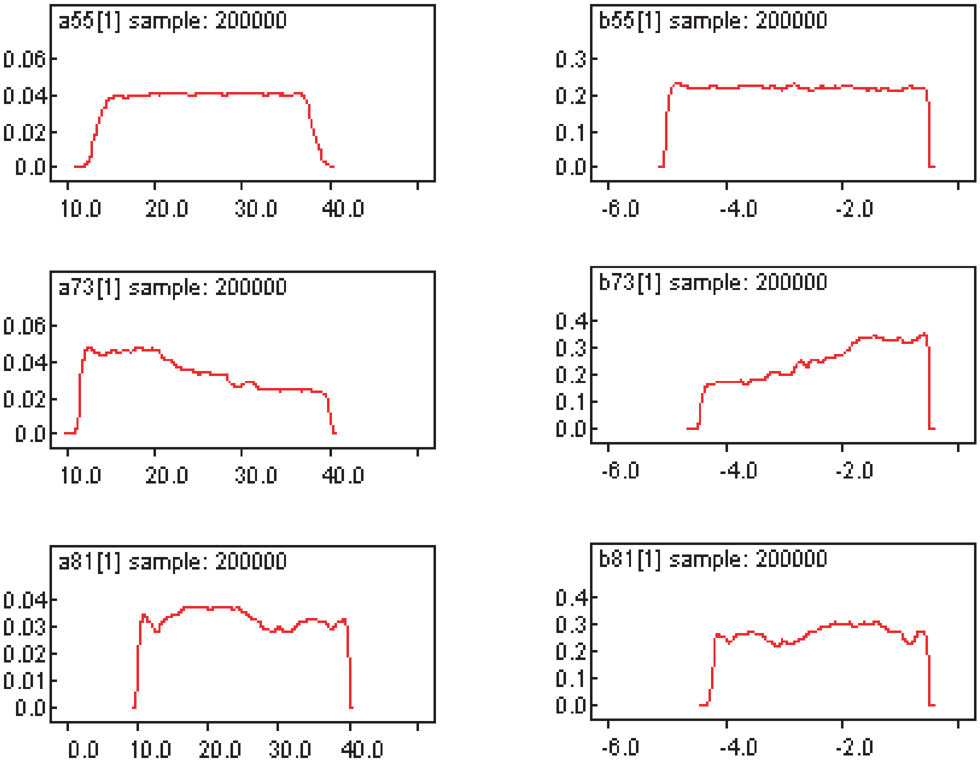


Figure 24: Posterior kernel densities of some a_i's and b_i's of Function II under the prior Case I

The posterior statistics, including the mean, median, 2.5%, and 97.5% quantiles, of a_i's versus the nine accelerated electric fields from 5.5 to 8.1MV/cm are plotted in Figure 25. For comparison, the posterior mean, 2.5%, and 97.5% quantiles, of a from Function I are also shown in the same figure. Similarly, the posterior statistics of b_i's and b are plotted in Figure 26. Both figures show that, within the range of the electric field applied in the ALT, no obvious trend exists in the statistics of a_i's or b_i's, except the 2.5% quantile of b_i's. Therefore, it is reasonable to assume that the acceleration is identical over this range of electric fields. But to use Function I as the acceleration model requires the assumption that it is still valid at low electric fields.

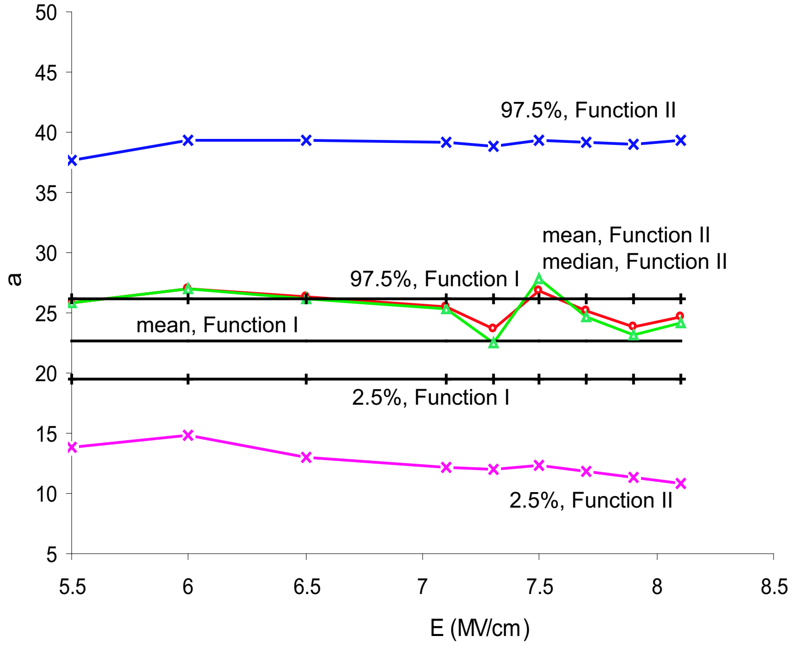


Figure 25: Comparison of posterior statistics under the prior Case I : a from Function I and a_i's from Function II

VI.5. Reliability projection

The Weibull scale parameter α under normal operating conditions or at a low electric field can be extrapolated with the acceleration function, i.e., $\ln \alpha_{\text{op}} = a + bE_{\text{op}} + \delta_{\text{op}}$, once a and b are estimated. The distribution of $\ln \alpha_{\text{op}}$ is determined from the

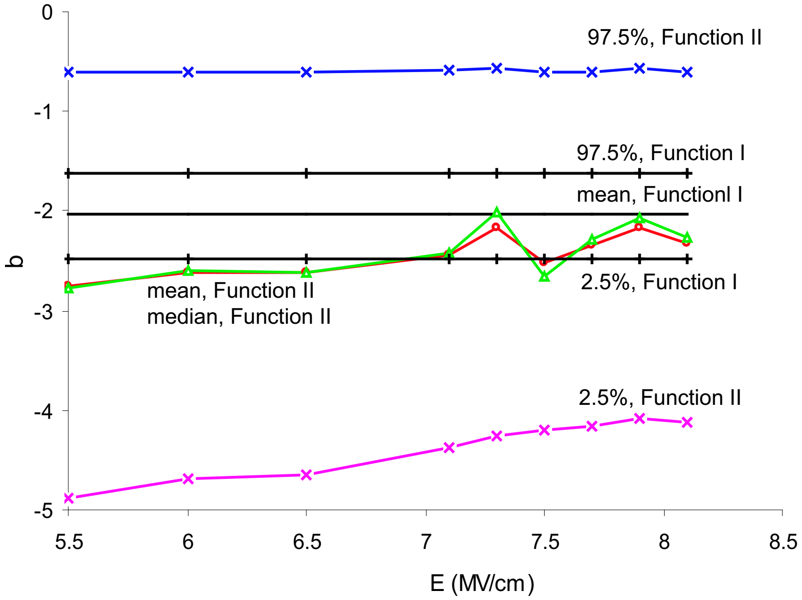


Figure 26: Comparison of posterior statistics under the prior Case I: b from Function I and b_i's from Function II

distributions of a and b, and δ_{op}. According to the graphic model shown in Figure 18, the distribution of α_{op} can be obtained directly in the WinBUGS simulation from the posteriors of a and b, if the distribution of δ_{op} is known.

For accuracy and convenience of computation, $\ln\alpha$ is used in the simulation instead of α. The simulation outputs of $\ln\alpha$ at low electric fields of 1.5MV/cm, 3.0MV/cm and 4.5MV/cm, together with the $\ln\alpha$'s at the nine accelerated stress levels, are given in Table 11. The posterior statistics are obtained under the prior Case II and the assumption of $\delta_{15}\sim\text{N}(0,1/49)$, $\delta_{30}\sim\text{N}(0,1/64)$, and $\delta_{45}\sim\text{N}(0,1/100)$, where δ_{15}, δ_{30}, and δ_{45} are the error terms at 1.5MV/cm, 3.0MV/cm, and 4.5MV/cm, respectively. The linear relationship between $\ln\alpha$ and E is plotted in Figure 27, together with the 95% credible bound. Normal priors with various variances have been assumed on δ_{15}, δ_{30} and δ_{45}. However, the difference is insignificant in the posterior statistics of $\ln\alpha$'s.

Table 11: Statistics of $\ln\alpha$ from simulation under the prior Case II, $\delta_{15}\sim\mathrm{N}(0,1/49)$, $\delta_{30}\sim\mathrm{N}(0,1/64)$, and $\delta_{45}\sim\mathrm{N}(0,1/100)$

Node[†]	Mean	Std.D.	MC error	2.5%	Median	97.5%
$\ln\alpha_{15}$	19.62	1.359	3.946×10^{-2}	17.07	19.59	22.38
$\ln\alpha_{30}$	16.58	1.023	2.961×10^{-2}	14.67	16.56	18.66
$\ln\alpha_{45}$	13.54	0.689	1.975×10^{-2}	12.26	13.53	14.95
$\ln\alpha_{55}$	11.52	0.464	1.317×10^{-2}	10.65	11.51	12.46
$\ln\alpha_{60}$	10.51	0.354	9.824×10^{-3}	9.849	10.51	11.23
$\ln\alpha_{65}$	9.495	0.254	6.600×10^{-3}	9.014	9.489	10.01
$\ln\alpha_{71}$	8.280	0.143	2.656×10^{-3}	8.005	8.279	8.564
$\ln\alpha_{73}$	7.877	0.121	1.376×10^{-3}	7.639	7.877	8.115
$\ln\alpha_{75}$	7.480	0.113	2.681×10^{-4}	7.257	7.481	7.701
$\ln\alpha_{77}$	7.070	0.122	1.259×10^{-3}	6.829	7.071	7.309
$\ln\alpha_{79}$	6.665	0.145	2.545×10^{-3}	6.379	6.665	6.947
$\ln\alpha_{81}$	6.254	0.176	3.844×10^{-3}	5.906	6.255	6.596

[†] As an example, $\ln\alpha_{55}$ denotes $\ln\alpha$ at $E = 5.5\mathrm{MV/cm}$.

Table 12 gives α calculated from the values of $\ln\alpha$ in Table 11. Because the kernel density of $\ln\alpha$ is very close to a normal density, the posterior mean is selected as the point estimator of $\ln\alpha$. The estimate of α corresponding to the mean of $\ln\alpha$ can be found in the columns titled "mean" of Table 12. They are

$$\begin{aligned}\hat{\alpha}_{15} &= 3.318\times10^{8}\text{ seconds or }3840.113\text{ days},\\ \hat{\alpha}_{30} &= 1.587\times10^{7}\text{ seconds or }183.691\text{ days},\\ \hat{\alpha}_{45} &= 7.592\times10^{5}\text{ seconds or }8.787\text{ days}.\end{aligned}$$

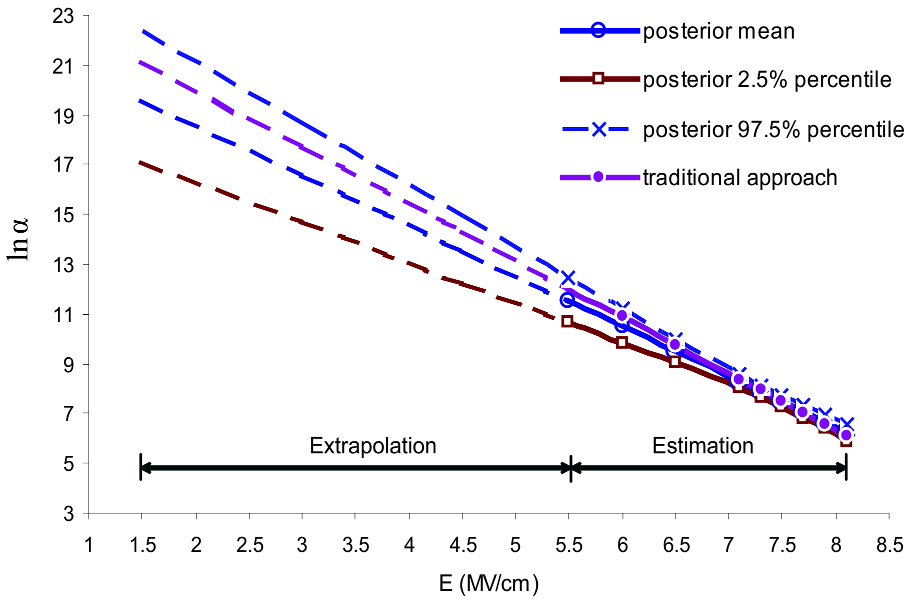


Figure 27: $\ln\alpha$ versus E plotted from simulation under the Case II prior, $\delta_{15}\sim\mathrm{N}(0,1/49)$, $\delta_{30}\sim\mathrm{N}(0,1/64)$, and $\delta_{45}\sim\mathrm{N}(0,1/100)$

In terms of α, the start and end points of the range that corresponds to the 95% credible interval of $\ln\alpha$ can be found in the respective columns titled "2.5%" and "97.5%" of Table 12. Next, take the posterior mean as the estimator of β. Table 9 gives $\hat{\beta} = 0.667$ and the 95% credible interval of (0.597,0.740). Then, the lifetime distributions at 1.5MV/cm, 3.0MV/cm, and 4.5MV/cm are estimated as $\mathrm{W}(3.318\times10^8, 0.667)$, $\mathrm{W}(1.587\times10^7, 0.667)$, and $\mathrm{W}(7.592\times10^5, 0.667)$, respectively.

VI.6. Comparison of two projections

Projecting the reliability at the operating conditions with the traditional approach can be done in the following three steps.

1. Fit the time-to-breakdown data at each electric field from 7.1 to 8.1MV/cm with a Weibull distribution and find the maximum likelihood estimate (MLE) of the Weibull parameters α and β. The results can be found in Table 13.

Table 12: α calculated from the simulation results of $\ln \alpha$ in Table 11

	α in second			α in day		
	Mean†	2.5%‡	97.5%$^\sharp$	Mean†	2.5%‡	97.5%$^\sharp$
α_{15}	3.318×10^{8}	2.591×10^{7}	5.242×10^{9}	3.840×10^{3}	2.998×10^{2}	6.067×10^{4}
α_{30}	1.587×10^{7}	2.350×10^{6}	1.270×10^{8}	1.837×10^{2}	2.720×10^{1}	1.470×10^{3}
α_{45}	7.592×10^{5}	2.111×10^{5}	3.110×10^{6}	8.787×10^{0}	2.443×10^{0}	3.599×10^{1}
α_{55}	1.007×10^{5}	4.219×10^{4}	2.578×10^{5}	1.166×10^{0}	4.883×10^{-1}	2.984×10^{0}
α_{60}	3.668×10^{4}	1.894×10^{4}	7.536×10^{4}	4.245×10^{-1}	2.192×10^{-1}	8.722×10^{-1}
α_{65}	1.329×10^{4}	8.217×10^{3}	2.225×10^{4}	1.539×10^{-1}	9.511×10^{-2}	2.575×10^{-1}
α_{71}	3.944×10^{3}	2.996×10^{3}	5.240×10^{3}	4.565×10^{-2}	3.467×10^{-2}	6.064×10^{-2}
α_{73}	2.636×10^{3}	2.078×10^{3}	3.344×10^{3}	3.051×10^{-2}	2.405×10^{-2}	3.871×10^{-2}
α_{75}	1.772×10^{3}	1.418×10^{3}	2.211×10^{3}	2.051×10^{-2}	1.641×10^{-2}	2.559×10^{-2}
α_{77}	1.176×10^{3}	9.243×10^{2}	1.494×10^{3}	1.361×10^{-2}	1.070×10^{-2}	1.729×10^{-2}
α_{79}	7.845×10^{2}	5.893×10^{2}	1.040×10^{3}	9.079×10^{-3}	6.821×10^{-3}	1.204×10^{-2}
α_{81}	5.201×10^{2}	3.672×10^{2}	7.322×10^{2}	6.020×10^{-3}	4.250×10^{-3}	8.474×10^{-3}

$^\dagger\alpha$ calculated from the mean of $\ln \alpha$ in Table 11.

$^\ddagger\alpha$ calculated from the 2.5% quantile of $\ln \alpha$ in Table 11.

$^\sharp\alpha$ calculated from the 97.5% quantile of $\ln \alpha$ in Table 11.

Failure data at 5.5 to 6.5MV/cm are not used due to small sample size.

2. Use linear regression on $\ln \hat{\alpha}_{\mathrm{MLE}}$'s at 7.1 to 8.1MV/cm to find the estimates of a and b of the acceleration function $\ln \alpha_i = a + bE_i$. This gives $\hat{a} = 24.535$, $\hat{b} = -2.275$ as shown in Figure 28.

3. Use $\widehat{\ln \alpha} = 24.535 - 2.275E$ to extrapolate $\ln \alpha$ at low electric fields. Given E, $\hat{\alpha} = e^{24.535-2.275E}$ is taken as the estimate of α. Average $\hat{\beta}_{\mathrm{MLE}}$'s in Table 13 to get $\hat{\beta} = 0.679$. The results of this step can be found in the middle of Table 14.

Table 13: MLE of Weibull parameters from 7.1 to 8.1MV/cm

E (MV/cm)	7.1	7.3	7.5	7.7	7.9	8.1
$\hat{\alpha}_{\mathrm{MLE}}$	3355	2451	2853	1302	736.1	334.2
$\hat{\beta}_{\mathrm{MLE}}$	0.683	0.655	0.879	0.668	0.578	0.610

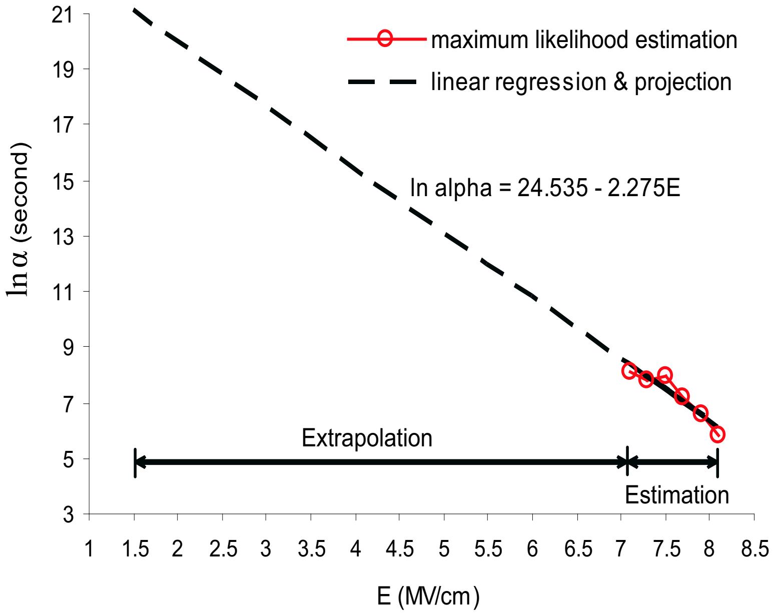


Figure 28: Linear regression of $\ln \hat{\alpha}_{\mathrm{MLE}}$ versus E and extrapolation to low electric fields

Table 14: Comparison of the estimation and projection between traditional and Bayesian approaches

E	Traditional Approach			Bayesian Approach		
(MV/cm)	$\widehat{\ln\alpha}$	$\hat{\alpha}$ (sec.)	$\hat{\alpha}$ (day)	$\widehat{\ln\alpha}$	$\hat{\alpha}$ (sec.)	$\hat{\alpha}$ (day)
1.5	21.12	1.490×10^{9}	1.725×10^{4}	19.62	3.318×10^{8}	3.840×10^{3}
3.0	17.71	4.912×10^{7}	5.685×10^{2}	16.58	1.587×10^{7}	1.837×10^{2}
4.5	14.30	1.619×10^{6}	1.873×10^{1}	13.54	7.592×10^{5}	8.787×10^{0}
5.5	12.02	1.664×10^{5}	1.926×10^{0}	11.52	1.007×10^{5}	1.166×10^{0}
6.0	10.88	5.334×10^{4}	6.173×10^{-1}	10.51	3.668×10^{4}	4.245×10^{-1}
6.5	9.747	1.710×10^{4}	1.979×10^{-1}	9.495	1.329×10^{4}	1.539×10^{-1}
7.1	8.382	4.367×10^{3}	5.054×10^{-2}	8.280	3.944×10^{3}	4.565×10^{-2}
7.3	7.927	2.770×10^{3}	3.207×10^{-2}	7.877	2.636×10^{3}	3.051×10^{-2}
7.5	7.472	1.758×10^{3}	2.034×10^{-2}	7.480	1.772×10^{3}	2.051×10^{-2}
7.7	7.017	1.115×10^{3}	1.291×10^{-2}	7.070	1.176×10^{3}	1.361×10^{-2}
7.9	6.562	7.075×10^{2}	8.188×10^{-3}	6.665	7.845×10^{2}	9.079×10^{-3}
8.1	6.107	4.489×10^{2}	5.195×10^{-3}	6.254	5.201×10^{2}	6.020×10^{-3}
	$\hat{a} = 24.535,\ \hat{b} = -2.275$ $\hat{\beta} = 0.679$			$\hat{a} = 22.653,\ \hat{b} = -2.024$ $\hat{\beta} = 0.667$		

The results obtained with the traditional and Bayesian approaches are compared in Table 14 and Figure 27. The two estimates of β are very close. This can be taken as additional evidence that the sample data have an identical failure mechanism. However, due to the discrepancy in both $\hat{a}$ and $\hat{b}$ between the two approaches, the difference in $\widehat{\ln \alpha}$ and hence $\hat{\alpha}$ are significant. Furthermore, the difference is amplified by extrapolation to low electric fields. The traditional projection gives relatively optimistic reliability at a low stress, but it is within the 95% credible bound of the Bayesian projection.

Unfortunately, because failure data under the operating conditions are not available in large quantities, it is impossible to verify the extrapolation and to judge which of the two projections better represents reality. Nevertheless, it is important to recognize that these are two different approaches that one can choose from to project reliability or use both as references for decision-making.

Two advantages should be emphasized for the Bayesian approach to lifetime projection. Firstly, the Bayesian approach is applicable when the sample size is small. This is very useful nowadays as it is generally either too costly or impossible to obtain a large sample of failure data due to increasing product reliability and time limitation. However, a relatively large sample size is required for the traditional approach in which the maximum likelihood estimator is used. This results in the small sample size of breakdown data collected at relatively low electric fields of 5.5 to 6.5MV/cm cannot be included in the traditional projection process.

Secondly, the Bayesian approach allows the construction of a broad class of estimators from the posterior distribution, while the traditional approach gives just one point estimator. In the Bayesian context, an estimator is chosen to minimize the risk with respect to a loss function that is defined for the problem at hand. For simplicity, the Bayes estimator is used as an example for the reliability projection in this chapter. However, a different estimator can be derived if another loss function is defined.

CHAPTER VII

FAILURE RATE ESTIMATION

VII.1. Bathtub failure rate and the change point

In the life cycle of a semiconductor device, the failure rate usually has a bathtub shape as shown in Figure 29 [28]. This curve is the combined result of three failure rates: a decreasing failure rate (DFR) due to infant mortality, a constant failure rate (CFR) representing the useful lifetime, and an increasing failure rate (IFR) reflecting the wear-out or aging effect.

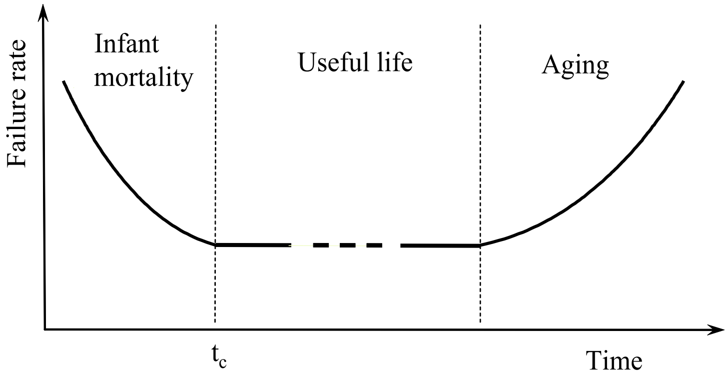


Figure 29: Bathtub failure rate curve

The high failure rate of infant mortality causes a large number of failures in the early stage of usage. To reduce or eliminate this kind of failure occurring on a customer's site, manufacturers employ a burn-in process to screen out the defective units before shipping [28]. The ultimate goal of conducting burn-in is to minimize the total cost of manufacturing, burn-in and warranty. Estimating the failure rate in the left hand side of the bathtub curve is the key to determining the optimal stress level and stress time of burn-in. Under the burn-in context, a product's infant mortality and useful life phases are of interest, while the wear-out phase is not the concern here.

The time when the failure rate turns from DFR into CFR is called the change point, denoted as t_c. In the literature, Weibull and exponential distributions have been used to describe the failure rate's behavior before and after t_c, respectively [2, 7]. The change point is usually estimated under the assumption that the function or the function form of DFR is known [5, 7]. The problem of simultaneously estimating t_c and the DFR requires considerable effort and remains an area for research.

Unlike the lifetime distribution discussed in the previous chapters, failure rate is used in this chapter to depict reliability. Different from the Weibull-exponential model or other parametric functions of failure rate assumed in the literature, no assumption is made on the function form of the infant mortality failure rate as long as it is a DFR.

The objective of this chapter is to provide a non-parametric procedure that can be used to estimate the DFR when the change point is unknown and the sample size is not large. To avoid the difficulty of estimating the DFR and t_c at the same time, two steps are taken on an x-set of n failure times placed in ascending order. At first, an exponentiality test is employed in determining the order j at which the sample is partitioned into two parts, N1=$(x_{(1)}, \cdots, x_{(j-1)})$ and N2=$(x_{(j)}, \cdots, x_{(n)})$, where N2 is tested for exponentiality against the alternatives of DFR and IFR. After that, a non-parametric Bayesian approach is applied to estimate the DFR of N1.

A rough failure rate estimate of the high k thin films at two electric fields are plotted in Figure 30. The failure rates are estimated with a non-parametric procedure designed for a relatively small sample size [41]. A single observation can have considerable influence on the result given by the procedure. However, the shape of the failure rate estimate is often used to suggest an appropriate model for the data.

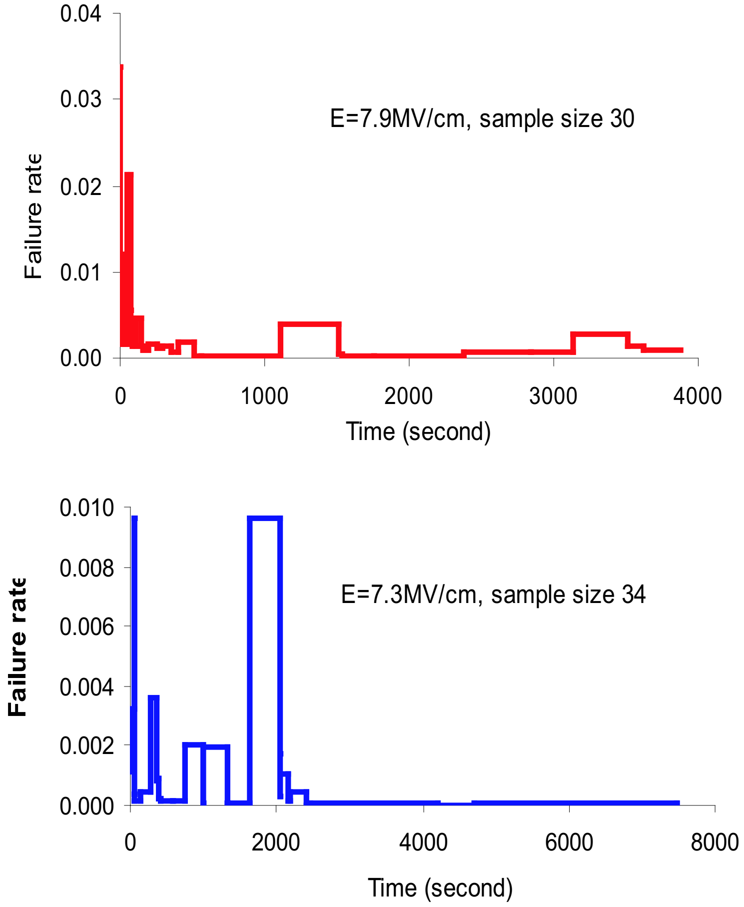


Figure 30: Empirical failure rate at E=7.9MV/cm and E=7.3MV/cm

From visual inspection, both plots demonstrate the pattern of a DFR followed by a CFR, where the change point t_c exists but is unknown. There is no sign of IFR due to the lack of failure data in the wear-out phase. For illustration purposes, the proposed procedure is applied on the ordered failure data at E=7.9MV/cm and E=7.3MV/cm given in Table 3 and Table 4.

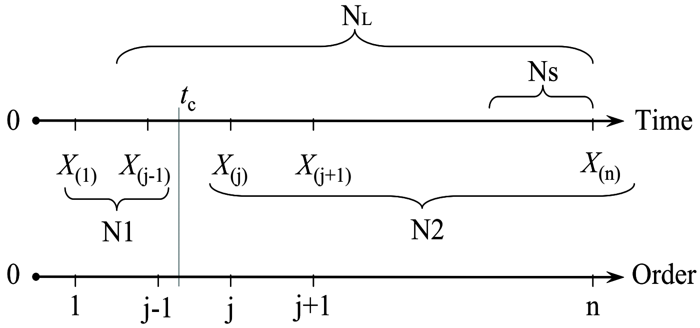


Figure 31: Illustration of the partition of an ordered data set

VII.2. Test for exponentiality

The idea of using the exponentiality test for sample partition can be illustrated with Figure 31. The ordered data set of failure time $x_{(1)}, \cdots, x_{(n)}$ consists of two sections, the early N1 of DFR and the latter N2 of CFR. Because of the restriction in testing time and the nature of relatively slow aging effect, the number of data points that belong to the IFR phase is small or none depending on the stress level. For a small sample of IFR, the increasing trend still may not be observed if the sample is pooled together with the relatively large sample of CFR. However, the increasing trend can show up if the sample is mixed with a small number of data points that are located at the end of CFR.

Unfortunately, t_c is unknown, so is the sample point $x_{(j)}$ that partitions the ordered data set into N1 and N2. In case of N_L that the last several data points of DFR are wrongly taken as from CFR, N_L shows a DFR instead of the supposed CFR. N_S is the opposite case that a few data points of CFR are mistaken as from DFR, then N_S has a CFR or possibly IFR. Theoretically, a reasonable approximation of t_c for a given data set should be the sample point $x_{(j)}$ that leads to the N2 fitting well with an exponential distribution; if there are multiple attempted partition points that result in good fitting of N2, then the point of a lower or the lowest order is recommended.

VII.2.1. Data transformation

The general form of the density function of an exponential distribution $\mathrm{E}(\alpha, \beta)$ is

$$f(x; \alpha, \beta) = \frac{1}{\beta} \exp(-\frac{x - \alpha}{\beta}), \quad x \geq \alpha, \tag{7.1}$$

where α and β are the location and scale parameters, respectively. If $x_{(i)}$'s follow $\mathrm{E}(\alpha, \beta)$, where α and β are unknown, then the spacing $x_{(i+1)} - x_{(i)}$, i=1,$\cdots$,n−1, is a sample from $\mathrm{E}(0, \beta)$ [56]. Testing the goodness-of-fit of the x-set on $\mathrm{E}(\alpha, \beta)$ is equivalent to testing the goodness-of-fit of the spacings on $\mathrm{E}(0, \beta)$. Therefore, the null hypothesis H_0 that needs to be tested is that a sample fits $\mathrm{E}(0, \beta)$.

A large number of test procedures have been given for H_0 [56]. These tests can be classified into two broad groups, tests using the original data and tests based on the data after certain transformations. The two transformation that are often made on the lifetime data under the test of H_0 are the $\mathbb{N}$ transformation and the $\mathbb{K}$ transformation.

Now Suppose $x_{(i)}$, $i = 1, \cdots, n$, is an ordered sample from $\mathrm{E}(0, \beta)$. Define the spacing by

$$D_i = x_{(i+1)} - x_{(i)}, \quad i = 1, \cdots, n,$$

with $x_{(0)} = 0$. The normalized spacing x'_i, defined by

$$x'_i = (n + 1 - i)D_i, \quad i = 1, \cdots, n,$$

is independently and identically distributed as $\mathrm{E}(0, \beta)$. The transformation from the x-set to the x'-set is called the $\mathbb{N}$ transformation. The total time on test is calculated as

$$T_n = \sum_{i=1}^{n} x_{(i)} \quad \text{or} \quad T_n = T'_n = \sum_{i=1}^{n} x'_{(i)}.$$

Let

$$T'_i = \sum_{k=1}^{i} x'_{(k)} \quad \text{and} \quad y_{(i)} = \frac{T'_i}{T_n}, \quad i = 1, \cdots, n-1.$$

Then, the y-set is an ordered sample of size $n-1$ from the uniform distribution U(0,1). The transformation from the x-set to the y-set is called the $\mathbb{K}$ transformation.

There are good reasons for making the data transformation, particularly when the original data are lifetimes. Firstly, both the $\mathbb{N}$ and $\mathbb{K}$ transformations are invariant under the labelling of an unordered x-set. This is usually considered as a desirable property for a test procedure. Secondly, the unknown parameter β has been eliminated in the $\mathbb{K}$ transformation at the cost of losing one degree of freedom. More importantly, if the x-set does not follow an exponential distribution but has an IFR, the x'-set is stochastically decreasing with i; if the x-set has a DFR, the x'-set is stochastically increasing with i. Consequently, the pattern of the y-set reveals information about the alternative hypothesis, namely, $y_{(i)}$ moves toward one for IFR samples and toward zero for DFR samples.

VII.2.2. Test techniques

Tests for exponentiality that can be applied on the original x-set include χ^2 tests, empirical distribution function (EDF) tests, regression tests, and the tests based on sample moments [51, 55, 56]. After the $\mathbb{N}$ transformation, these tests may equally be applied to the x'-set. With the $\mathbb{K}$ transformation, the exponentiality test on the x-set can be dealt with as testing the uniformity of the y-set. This provides test procedures that generally have more power than those mentioned above.

The simplest test for the uniformity of the y-set is based on $\bar{Y}$, the mean of the y-set [34]. Let S be the sum of $y_{(i)}$, i=1,$\cdots$,$n-1$, then S and $\bar{Y}$ can be calculated from the x-set as follows.

$$S = 2n - 2\sum_{i=1}^{n} \frac{ix_{(i)}}{T_n} = 2\sum_{i=1}^{n} \frac{(n-i)x_{(i)}}{T_n},$$

$$\bar{Y} = \frac{S}{n-1}.$$

The distribution of $\bar{Y}$ converges quickly to the normal. Specifically, $\bar{Y}$ follows approximately the normal distribution $\mathrm{N}(0.5, 1/12m)$, where $m = n-1$ ($m \geq 15$).

Because $\bar{Y}$ tends to be large for an original IFR sample and to be small for a DFR sample, $\bar{Y}$ can be used as a one-tail statistic to guard against the alternative of either IFR or DFR. But it will be two-tailed against unknown or general alternatives. A power study on the various test procedures for H_0 shows that $\bar{Y}$, along with two of the EDF statistics, the Anderson-Darling statistic A^2 and the Cramér-von statistic W^2, has a very high power in two-tail tests against both IFR and DFR alternatives. In one-tail tests that the direction of the alternative is known , $\bar{Y}$ is also effective and overall better than A^2. Based on the discussions of this section, $\bar{Y}$ is selected as the test statistic for the current problem.

VII.2.3. Sample partition

The plot of the empirical failure rate in Figure 30 gives a rough idea of the location where the DFR changes to the CFR. For example, at E=7.9MV/cm, the change occurs somewhere within the interval [150,300]. Therefore, the candidates of the partition point are from $x_{(12)} = 153$ to $x_{(15)} = 290$. In order to select one out of the four candidates, the $\mathbb{K}$ transformation is applied on the sample N2 given by each of the attempted partition point $x_{(i)}$, $i = 12, \cdots, 15$, and the statistic $\bar{Y}$ is computed to test the exponentiality. It is found that $x_{(13)} = 193$ results in the $\bar{Y}$ that is the closest to 0.5. This can be understood as the sample $(x_{(13)}, \cdots, x_{(32)})$ fits an exponential

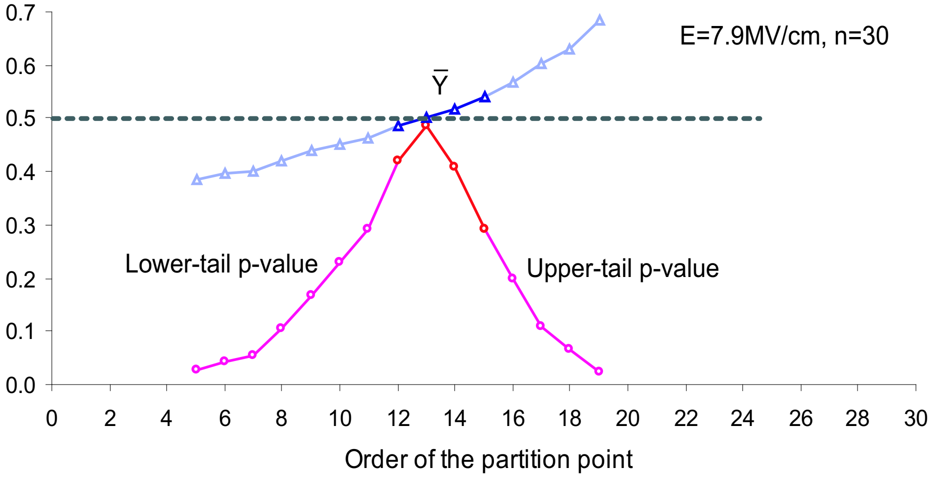


Figure 32: Test statistics versus the order of attempted partition point at E=7.9MV/cm

distribution best among all the possible choices of N2 from the x-set.

In fact, the statistic $\bar{Y}$ increases with the order of the attempted partition point. This is clearly seen in Figure 32 where the corresponding p-value is also shown. This trend is consistent with the previous discussion that $\bar{Y}$ tends to be small for an original DFR sample and to be large for an IFR sample. As $\bar{Y}$ follows approximately a normal distribution with the mean of 0.5, a value close to 0.5 indicates it is more likely that N2 has a CFR rather than a DFR or IFR. If the attempted partition point is actually much earlier than t_c, this is the case of N$_L$, which has a DFR as shown in Figure 31. If the attempted partition point is much later than t_c, this falls into the case of N$_S$, which has a CFR or, possibly, an IFR if the wear-out starts before the end of the stress test due to a relatively high stress level. As mentioned early in the beginning of Section VII.2, the number of data points from the IFR phrase is very small in comparison with that from the CFR period and the increasing amount of failure rate is negligible. As a result, there is no sign of IFR in the failure rate plot of Figure 30. It is still reasonable to pool these IFR points with those of the CFR.

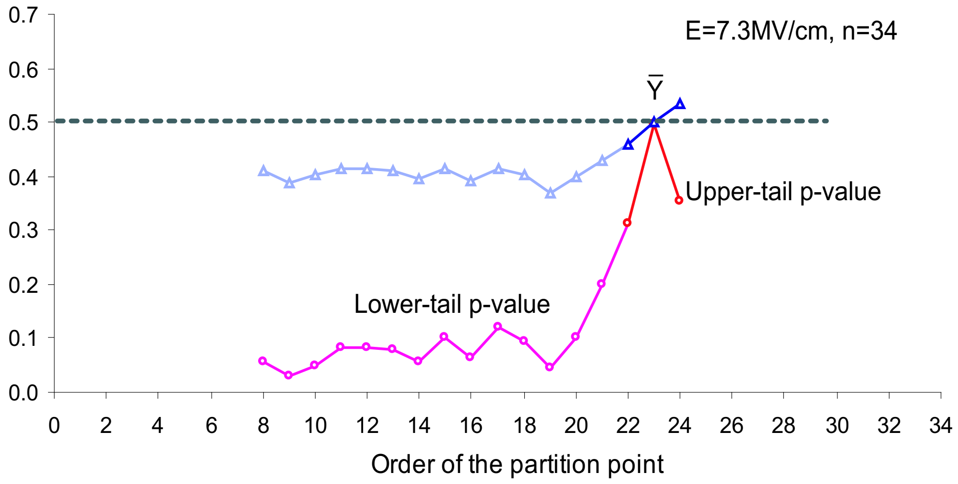


Figure 33: Test statistics versus the order of attempted partition point at E=7.3MV/cm

A similar result is obtained at E=7.3MV/cm. From the empirical failure rate curve in Figure 30, $x_{(22)}$ to $x_{(24)}$ are taken as the candidates to partition the sample of size 34. Among the three candidates, $x_{(23)}$=2181 leads to the $\bar{Y}$ that is the closest to 0.5. Figure 33 shows how $\bar{Y}$ and the p-value change with the order of the attempted partition point. $\bar{Y}$ stays at a low level when the attempted partition point is in the early part of the x-set; the increasing trend of $\bar{Y}$ does not start until the attempted point is larger than $x_{(19)}$. The cause of the early small $\bar{Y}$ is that, due to the relatively low stress field compared to E=7.9MV/cm, the infant mortality phase lasts longer before the period of CFR starts.

VII.3. Estimation of decreasing failure rate

After the partition point $x_{(j)}$ is determined, the sample N1=$(x_{(1)}, \cdots, x_{(j-1)})$ is used to estimate the DFR with the non-parametric Bayesian approach developed by Mazzuchi and Singpurwalla [43] and later implemented on burn-in by Chien and Kuo [6].

The approach is non-parametric in the sense that no assumption is made on the parametric form of the lifetime distribution except that it has a monotone failure rate.

VII.3.1. The non-parametric Bayesian approach

Let x be the random variable of the lifetime described by a distribution F, whose monotone failure rate in the time period $(0,\, t_h]$ is to be estimated. Let $\tau_1, \cdots, \tau_k$ be equally spaced points of time, where $0 \equiv \tau_0 < \tau_1 < \cdots < \tau_k \equiv t_h$ and $\Delta\tau$ is the length of each interval. Also let h_i be the interval failure rate in $(\tau_{i-1}, \tau_i]$. The goal is to approximate the monotone failure rate with the interval failure rate h_i's.

First define u_i as the probability that a unit will not fail in $(\tau_{i-1}, \tau_i]$ conditioning on that it has survived in $(0, \tau_{i-1}]$,

$$u_i = \Pr(x > \tau_i | x > \tau_{i-1}) = \frac{1 - F(\tau_i)}{1 - F(\tau_{i-1})}, \quad i = 1, \cdots, k, \tag{7.2}$$

and $u_0 = 0$, $u_{k+1} = 1$. Because $F(t) = 1 - e^{-H(t)}$, where $H(t)$ is the hazard function, this gives the relationship of $u_i \approx e^{-h_i \Delta\tau}$. Then, the interval failure rate h_i can be calculated from [6]

$$h_i \approx -\frac{\ln u_i}{\Delta\tau}. \tag{7.3}$$

Now the problem becomes finding an estimation of the u_i's. A Bayesian approach has been developed for this end.

Let s_i denote the number of breakdown events that are observed in $(\tau_{i-1}, \tau_i]$. Given $\mathbf{s} = (s_1, \cdots, s_k)$, the likelihood of $\mathbf{u} = (u_1, \cdots, u_k)$ is

$$l(\mathbf{u}|\mathbf{s}) \propto \prod_{i=1}^{k} (1 - u_i)^{s_i} u_i^{v_i},$$

where $v_i = \sum_{j=i+1}^{k} s_j$ is the number of units that survive in $(\tau_{i-1}, \tau_i]$. In the case of DFR where $0 \leq u_1 \leq u_2 \leq \cdots \leq u_k$, assume the unknown quantities $u_1, (u_2 - u_1), (u_3 - u_2), \cdots, (u_k - u_{k-1})$, and $(1 - u_k)$ have a prior distribution of the Dirichlet density with parameters $\alpha_i, i = 1, \cdots, k+1$, and β,

$$\pi(\mathbf{u}) = \frac{\Gamma(\beta)}{\prod_{i=1}^{k+1} \Gamma(\beta\alpha_i)} \prod_{i=1}^{k+1} (u_i - u_{i-1})^{\beta\alpha_i - 1}, \tag{7.4}$$

where $\alpha_i \geq 0, i = 1, \cdots, k+1, \sum_{i=1}^{k+1} \alpha_i = 1$, and $\beta > 0$.

By the Bayes' rule, the posterior density of $\mathbf{u}$ is written as

$$\pi(\mathbf{u}|\mathbf{s}) = \frac{l(\mathbf{u}|\mathbf{s})\pi(\mathbf{u})}{\int_{\mathbf{u}} l(\mathbf{u}|\mathbf{s})\pi(\mathbf{u}) d\mathbf{u}} = \frac{[\prod_{i=1}^{k}(1-u_i)^{s_i} u_i^{v_i}][\prod_{i=1}^{k+1}(u_i - u_{i-1})^{\beta\alpha_i - 1}]}{\int_{\mathbf{u}} [\prod_{i=1}^{k}(1-u_i)^{s_i} u_i^{v_i}][\prod_{i=1}^{k+1}(u_i - u_{i-1})^{\beta\alpha_i - 1}] d\mathbf{u}}.$$

Conceptually, once $\pi(\mathbf{u}|\mathbf{s})$ is known, the interval failure rate h_i's can be estimated through the formula (7.3). If $\Delta\tau$ is small enough, then h_i's is a good approximation of the monotone failure rate.

VII.3.2. Prior parameters and posterior simulation

The first issue in implementing the non-parametric approach to the DFR estimation is to determine the prior, i.e., to assign an appropriate value to the Dirichlet parameters α_i's and β. According to Mazzuchi and Singpurwalla [43], the marginal expectation of u_i in the Dirichlet density (7.4) is

$$\mathbb{E}(u_i) = \sum_{1}^{i} \alpha_j.$$

Therefore, α_i can be set as $\alpha_i = u_i^* - u_{i-1}^*$, where u_i^* is the best prior guess of u_i, except that $u_0^* = 0, u_{k+1}^* = 1$. Also from

$$\text{Var}(u_i) = \frac{(\sum_{1}^{i} \alpha_j)(\sum_{i+1}^{k+1} \alpha_j)}{\beta + 1},$$

β is related to the variance of u_i. Hence, a larger β should be used to reflect the variance reduction of the analysis. Generally, the value of β is determined based on one's knowledge about the device. If the device characteristics are not well known, the suggested value for β is between 1 and 20 [6].

The second issue is how to obtain the posterior $\pi(\mathbf{u}|\mathbf{s})$. Because k, the number of $\Delta\tau$ intervals, is usually large, Markov chain Monte Carlo simulation is a good option to avoid the high-dimensional integration involved. By including the relationship of (7.3) into the WinBUGS model, the distribution of h_i's, denoted as $\pi(\mathbf{h}|\mathbf{s})$, can be obtained directly from the simulation. For the data set at one stress level, a variety of Dirichlet priors which have different values of α_i's and β are put into the simulation. In all cases, $\pi(h_i|\mathbf{s})$ approximates a normal density; hence the mean of $\pi(h_i|\mathbf{s})$ is chosen as the estimator of h_i.

For illustration, take the data set at E=7.9MV/cm as an example. With the partition point $x_{(13)} = 193$, the sample for DFR estimation is N1=$(x_{(1)}, \cdots, x_{(12)})$, where $x_{(12)} = 153$. Next, partition the time period of (0,153] into 10 intervals, each with the length of $\Delta\tau$=15.3 seconds. Table 15 gives four Dirichlet priors based on two sets of α_i's and two β values, so that one can examine the influence of α_i's and of β on the posterior. Identical values of α_i's in the priors a-1 and a-2 represent the belief that u_i's decrease the same amount in each interval, whereas the reducing values of α_i's in the priors b-1 and b-2 convey the idea that the decreasing of u_i slows down as time proceeds.

In the WinBUGS simulation, four Markov chains run simultaneously. Each chain continues for 6000 iterations, whereas the first 1000 iterations are discarded before any data analysis. The good convergence is confirmed by the fact that the German-Rubin statistic R is stable around 1. It is found that the kernel density of h_i is very much like a normal density. As an example, the output statistics of the

Table 15: Dirichlet priors used for the data set at E=7.9MV/cm

Prior	Dirichlet parameters											
	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	β
a-1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	8
a-2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	15
b-1	0.25	0.25	0.2	0.15	0.15	0.15	0.1	0.1	0.1	0.05	0.05	8
b-2	0.25	0.25	0.2	0.15	0.15	0.15	0.1	0.1	0.1	0.05	0.05	15

WinBUGS simulation under the prior a-1 is listed in Table 16. The values in the column of "Mean" and "St.D." can be used as the mean and standard deviation, respectively, of the normal distribution that $\pi(h_i|\mathbf{s})$ approximates.

The estimates of h_i's with the mean of $\pi(h_i|\mathbf{s})$ are plotted in Figure 34. Comparisons are made among the DFRs obtained from the four priors given in Table 15. When two priors have the same α_i's, such as a-1 versus a-2 and b-1 versus b-2, a smaller β which corresponds to a larger variance of u_i makes the early part of the estimated DFR lower than that from a larger β; but this relation switches in the latter part of the failure rate curve. Thus the prior with a small β tend to give a relatively low and flat DFR. In the case of a same β, such as a-1 versus b-1 and a-2 versus b-2, the prior with decreasing α_i's results in a higher DFR that drops quickly in the early time period. After that, it is consistently lower than the DFR estimated from the prior with equal α_i's, even though the difference diminishes as the time proceeds. Hence, for a sharp DFR, the Dirichlet prior should be set with decreasing α_i's. In addition, the larger the difference between consecutive α_i's, the faster the DFR decreases.

At E=7.3MV/cm, the sample for DFR estimation is N1=$(x_{(1)}, \cdots, x_{(22)})$, where $x_{(22)} = 2154$. Set $t_h = 2154$ and $\Delta\tau = 179.5$ seconds. The DFR is estimated

Table 16: Statistics of h_i's at E=7.9MV/cm from WinBUGS simulation with the prior a-1

Node	Mean (10^{-2})	Std.D. (10^{-2})	MC error (10^{-4})	2.5% (10^{-2})	Median (10^{-2})	97.5% (10^{-2})
h_1	7.89	2.49	2.79	4.19	7.49	13.8
h_2	5.62	1.52	1.52	3.12	5.46	9.04
h_3	4.83	1.28	1.27	2.69	4.71	7.65
h_4	4.27	1.16	1.06	2.32	4.17	6.79
h_5	3.82	1.08	0.96	1.98	3.72	6.21
h_6	3.34	1.02	0.89	1.63	3.24	5.58
h_7	2.86	0.96	0.91	1.26	2.77	5.01
h_8	2.34	0.91	0.90	0.85	2.24	4.39
h_9	1.87	0.84	0.82	0.55	1.77	3.78
h_{10}	1.44	0.75	0.67	0.31	1.32	3.18

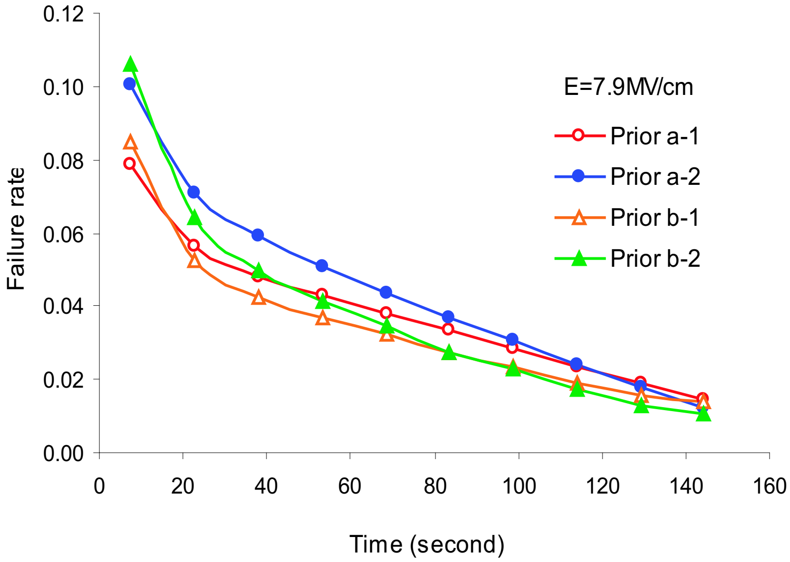


Figure 34: Estimated DFR at E=7.9MV/cm with the priors given in Table 15

with the priors given in Table 17, and the estimates are plotted in Figure 35. Comparisons made within the four failure rate curves lead to the same observation and conclusion as that from Figure 34. It is noticed that the failure rate at 7.3MV/cm is much lower that than that at 7.9MV/cm.

Table 17: Dirichlet priors used for the data set at E=7.3MV/cm

Prior	Dirichlet parameters													
	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}	α_{13}	β
a-1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	9
a-2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	15
b-1	0.3	0.2	0.2	0.15	0.15	0.15	0.15	0.1	0.1	0.1	0.1	0.05	0.05	9
b-2	0.3	0.2	0.2	0.15	0.15	0.15	0.15	0.1	0.1	0.1	0.1	0.05	0.05	15

VII.4. Estimation of constant failure rate

Assume that time-to-breakdown in the useful life period follows an exponential distribution E(α, β) with the density function of equation 7.1. The goal is to estimate the constant failure rate λ with the sample N2=$(x_{(j)}, \cdots, x_{(n)})$. Since $\lambda = \beta^{-1}$ and α is not related to λ, it is equivalent to consider estimating the failure rate of E$(0, \beta)$ with the sample $(x_{(j+1)} - x_{(j)}, \cdots, x_{(n)} - x_{(j)})$. In terms of λ, the density function of E$(0, \beta)$ is

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad \lambda > 0.$$

For simplification, write $n_s = n - j$ and

$$\mathbf{t} = (t_1, \cdots, t_{n_s}) = (x_{(j+1)} - x_{(j)}, \cdots, x_{(n)} - x_{(j)}).$$

The notation n_s and $\mathbf{t}$ will be used in the derivation that follows.

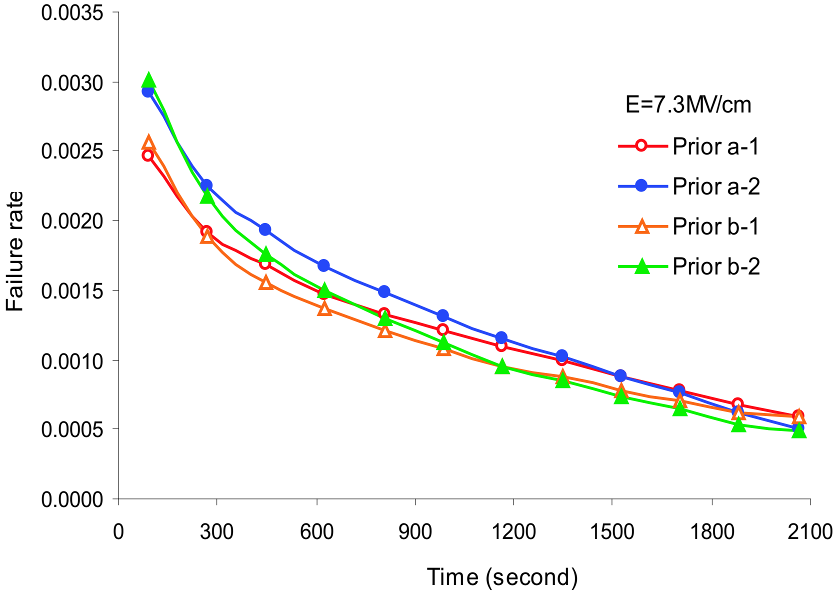


Figure 35: Estimated DFR at E=7.3MV/cm with the priors given in Table 17

Although there are many ways to estimate λ, a Bayesian approach is preferred here since the sample size of N2 may be small. In the Bayesian context, λ is treated as a random variable which is given a prior distribution $\pi(\lambda)$. Because little information is available regarding the useful life of high k dielectric thin films, $\pi(\lambda)$ is either non-informative or very vague as in the following two cases.

a. Jeffreys' prior

In the absence of prior information about an unknown parameter, Jeffreys' prior, which is non-informative and defined in terms of the Fisher information, can be used [25]. Jeffreys' prior is a uniform density on the space of probability distributions in the sense that it assigns equal mass to each different distribution. A unique feature of Jeffreys' prior is that it is invariant under parameter transformation.

For an exponential sampling distribution, the Jeffreys' prior on λ is proportional to λ^{-1} [41]. Simply let $\pi(\lambda) = \lambda^{-1}$, then the joint density of $(\lambda, \mathbf{t})$ is

$$f(\lambda, \mathbf{t}) = f(\mathbf{t}|\lambda)\pi(\lambda) = \lambda^{n_s-1} \exp(-\lambda \sum_{i=1}^{n_s} t_i),$$

and the marginal density of $\mathbf{t}$ is

$$f(\mathbf{t}) = \int f(\lambda, \mathbf{t}) d\lambda = \int \lambda^{n_s-1} \exp(-\lambda \sum_{i=1}^{n_s} t_i) d\lambda = \frac{(n_s - 1)!}{(\sum_{i=1}^{n_s} t_i)^{n_s}}.$$

This gives the posterior density of λ as

$$\pi(\lambda|\mathbf{t}) = \frac{f(\lambda, \mathbf{t})}{f(\mathbf{t})} = \frac{(\sum_{i=1}^{n_s} t_i)^{n_s}}{(n_s - 1)!} \lambda^{n_s-1} \exp(-\lambda \sum_{i=1}^{n_s} t_i).$$

Compare $\pi(\lambda|\mathbf{t})$ with a gamma density with the parameters α and β, i.e.,

$$g(\lambda; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \lambda^{\alpha-1} e^{-\lambda/\beta}.$$

It is clear that $\pi(\lambda|\mathbf{t})$ is a gamma density with the parameters $\alpha = n_s$ and $\beta = (\sum_{i=1}^{n_s} t_i)^{-1}$. Hence, the posterior $\pi(\lambda|\mathbf{t})$ is proper although the prior $\pi(\lambda) = \lambda^{-1}$ is not.

The posterior mean and variance are

$$\begin{aligned} \mathbb{E}(\lambda|\mathbf{t}) &= \alpha\beta = \frac{n_s}{\sum_{i=1}^{n_s} t_i}, \\ \mathrm{Var}(\lambda|\mathbf{t}) &= \alpha\beta^2 = \frac{n_s}{(\sum_{i=1}^{n_s} t_i)^2}. \end{aligned}$$

The Bayes estimator of λ with respect to the squared-error loss function is the posterior mean, i.e.,

$$\hat{\lambda} = \mathbb{E}(\lambda|\mathbf{t}) = \frac{n_s}{\sum_{i=1}^{n_s} t_i}.$$

Intuitively, $\hat{\lambda}$ is the reciprocal of the average time-to-breakdown of the n_s data points. This is equal to the maximum likelihood estimator which comes solely from the sampling data without any prior information. This is not coincidence at all if one realizes that Jeffreys' prior is non-informative.

b. Truncated uniform prior

An upper bound of λ can be found from the DFR estimated from the sample N1. Hence, the prior on λ can be assumed as a uniform density over an interval $(0, w)$,

$$\pi(\lambda) = \frac{1}{w}, \quad 0 < \lambda < w.$$

The joint density of $(\lambda, \mathbf{t})$, the marginal density of $\mathbf{t}$, and the posterior density of λ are derived in order as follows.

$$f(\lambda, \mathbf{t}) = f(\mathbf{t}|\lambda)\pi(\lambda) = \frac{\lambda^{n_s}}{w} \exp(-\lambda \sum_{i=1}^{n_s} t_i),$$

$$f(\mathbf{t}) = \int f(\lambda, \mathbf{t}) d\lambda = \frac{1}{w} \int_v^w \lambda^{n_s} \exp(-\lambda \sum_{i=1}^{n_s} t_i) d\lambda,$$

$$\pi(\lambda|\mathbf{t}) = \frac{f(\lambda, \mathbf{t})}{f(\mathbf{t})} = \frac{\lambda^{n_s} \exp(-\lambda \sum_{i=1}^{n_s} t_i)}{\int_0^w \lambda^{n_s} \exp(-\lambda \sum_{i=1}^{n_s} t_i) d\lambda}.$$

To calculate the integration in the denominator of $\pi(\lambda|\mathbf{t})$, write $\tau = \sum_{i=1}^{n_s} t_i$ and $z = \lambda\tau$. Then the denominator can be computed as

$$\int_0^w \lambda^{n_s} \exp(-\lambda \sum_{i=1}^{n_s} t_i) d\lambda = \int_0^{w\tau} \frac{z^{n_s} e^{-z} dz}{\tau^{n_s+1}} = \frac{\Gamma(n_s + 1, w\tau)}{\tau^{n_s+1}},$$

where $\Gamma(n_s + 1, w\tau)$ represents an incomplete gamma function. Therefore, the posterior can be written as

$$\pi(\lambda|\mathbf{t}) = \frac{\tau^{n_s+1} \lambda^{n_s} e^{-\lambda\tau}}{\Gamma(n_s + 1, w\tau)}.$$

The posterior mean and the second moment are computed to be

$$\mathbb{E}(\lambda|\mathbf{t}) = \frac{\Gamma(n_s+2, w\tau)}{\tau\Gamma(n_s+1, w\tau)},$$

$$\mathbb{E}(\lambda^2|\mathbf{t}) = \frac{\Gamma(n_s+3, w\tau)}{\tau^2\Gamma(n_s+1, w\tau)}.$$

The Bayes estimator of λ is the posterior mean $\mathbb{E}(\lambda|\mathbf{t})$, and the posterior variance can be found using the relationship of

$$\mathrm{Var}(\lambda|\mathbf{t}) = \mathbb{E}(\lambda^2|\mathbf{t}) - \mathbb{E}^2(\lambda|\mathbf{t}).$$

At the end of this section, the estimation of the CFR at E=7.9MV/cm and E=7.3MV/cm is summarized in Table 18. At each stress, the posteriors derived from Jeffreys' prior and the truncated uniform prior have the similar mean and variance. The difference in mean is less than 7%. This is probably because the priors are either non-informative or vague, hence the sample data has dominating influence on the posterior. If the sample size n_s decreases, the difference in posteriors between the two priors is expected to increase. Nevertheless, the estimation from the truncated uniform prior is the better one of the two.

Table 18: Estimation of CFR at E=7.9MV/cm and E=7.3MV/cm

	E=7.9MV/cm, $n_s = 17$		E=7.3MV/cm, $n_s = 11$	
$\pi(\lambda)$	Jeffreys' prior λ^{-1}	Uniform on $(0, 1.44\times10^{-2})$	Jeffreys' prior λ^{-1}	Uniform on $(0, 5.86\times10^{-4})$
$\pi(\lambda|\mathbf{t})$	$(3.33\times10^{62}) \cdot\lambda^{17}e^{-28932\lambda}$	$(5.67\times10^{65}) \cdot\lambda^{17}e^{-28932\lambda}$	$(4.11\times10^{46}) \cdot\lambda^{10}e^{-68234\lambda}$	$(2.55\times10^{50}) \cdot\lambda^{10}e^{-68234\lambda}$
$\mathbb{E}(\lambda|\mathbf{t})$	5.88×10^{-4}	6.22×10^{-4}	1.61×10^{-4}	1.76×10^{-4}
$\mathrm{Var}(\lambda|\mathbf{t})$	2.03×10^{-8}	2.15×10^{-8}	2.36×10^{-9}	2.58×10^{-9}

VII.5. Summary of the estimation

Because there is little prior information on the failure rate of high k thin films, vague priors are highly recommended. The DFR estimated from the Dirichlet prior with the parameters α_i=0.1, i=1,$\cdots$,k+1, and β=8 at E=7.9MV/cm or β=9 at E=7.3MV/cm, is chosen as the estimation of the true DFR; meanwhile, λ estimated from the truncated uniform prior is used as the estimate of the CFR. The two estimates are put together in Figure 36 to form the failure rate curve at each stress level.

Figure 36 shows that under a larger stress, the failure rate starts at a much higher level and decreases faster at the early stage; the DFR phase is shorter but the CFR is higher than that under a lower stress. From the failure rates estimated at different stress levels, one can study the acceleration behavior of the device and use this information to optimize the burn-in process. As burn-in is not the topic of the current study, no further discussion on this matter will be given here.

It should be pointed out that failure rate estimation in the literature with the Weibull-exponential model uses the fact that the DFR and the CFR are equal at the change point t_c; thus estimating t_c and DFR is sufficient. The approach presented in this chapter bypasses estimating t_c, but because the partition point generally is not the actual change point, the failure rate at the end of the estimated DFR curve differs with the CFR. Therefore, a better estimation of the CFR should be obtained from the sample N2 given by the partition point.

The maximum likelihood estimator of t_c in the Weibull-exponential model may not be unique [5]. It is difficult to determine the partition point $x_{(j)}$ with the exponentiality test if the statistic $\bar{Y}$ fluctuates around 0.5. A reasonably good sample of failure data will display the trend of $\bar{Y}$ increasing with j as shown in Figure 32

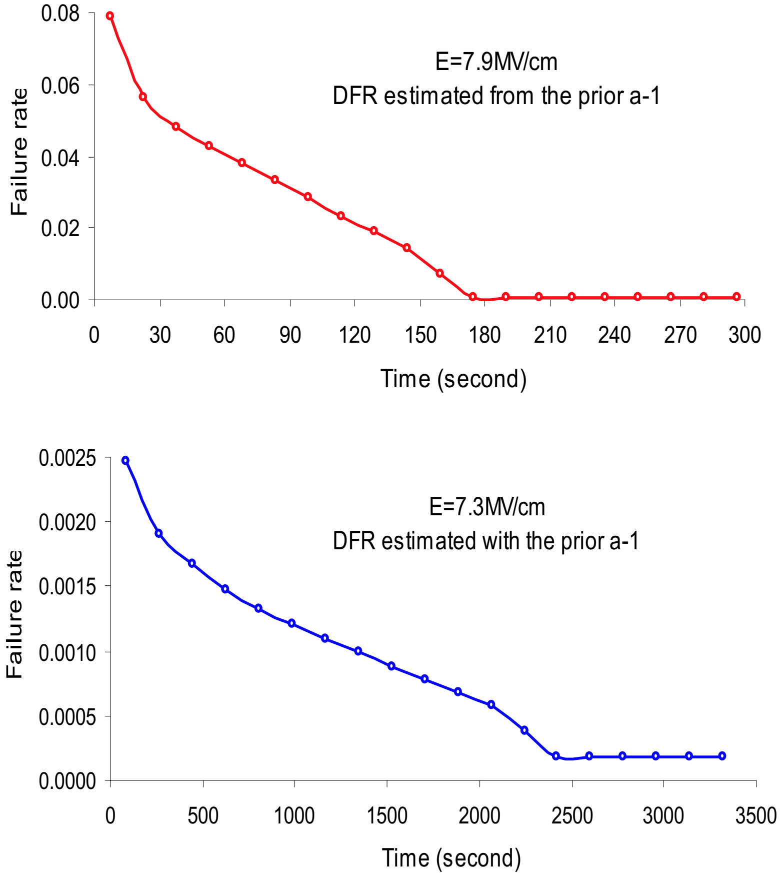


Figure 36: Estimated DFR and CFR at E=7.9MV/cm and E=7.3MV/cm

and Figure 33. Otherwise, an additional criterion or a different approach might need to be considered.

This approach to failure rate estimation has several advantages. It alleviates the difficulty of simultaneously estimating the change point and the failure rate by dividing it into two subproblems, each of which can be solved with established statistical methods and tools. In the literature, maximum likelihood estimation is usually used for the Weibull-exponential model; hence a large sample size is needed. But the approach given here is applicable even if sample size is relatively small, due to the Bayesian approach it takes to failure rate estimation. Furthermore, this makes it possible that failure rate information to be used for burn-in optimization can be obtained in a timely and efficient manner.

CHAPTER VIII

CONCLUSIONS

VIII.1. Summary

The topic of this dissertation is multidisciplinary, and the content is broad. It deals not only with the statistical aspects of reliability, but also with the electrical aspects of reliability characterization. As a special case, TaO_x-based high k dielectric thin films are studied in depth and analyzed with respect to failure mode identification, accelerated life testing, lifetime projection, and failure rate estimation. The results of this work can be summarized in three parts.

a. Failure mode identification and breakdown detection

A large number of experiments have been carried out on the ground of understandings of the methods and techniques for device characterization and stress testing. In Chapter III, a detailed study on dielectric relaxation and transient current presents strong experimental evidence from physics perspective that the relaxation current in high k dielectrics is different from the trapping/detrapping current of SiO_2 even though their behaviors look alike. High k thin films have a lower leakage current but a higher relaxation current than SiO_2. As a result, the relaxation current plays an important role in the degradation process and breakdown mode of high k films.

Due to the large relaxation current, dielectric relaxation is much more obvious in the high k films than in SiO_2. The connection between polarization-relaxation and film integrity is demonstrated in ramped voltage stress tests and explained from the materials and physics points of view. Because the relaxation property is essential to dielectrics, losing the ability to relax from a stress reflects the fact that the material is damaged or destroyed. Therefore, the disappearance of relaxation current can be used as the signal of a breakdown event, and this moment can be detected by monitoring the relaxation current. Because the signal on high k films is very sensitive, this method can catch the occurrence of a breakdown accurately and precisely. It is also easy to implement and does not need a quantitative failure criterion as most of the other detection methods do.

b. Projection of reliability under normal operating conditions

The projection model proposed in Chapter V combines the Weibull lifetime distribution with the empirical acceleration function and puts the model parameters into a hierarchical Bayesian framework. The essential idea of the Bayesian approach is that each model parameter is considered as a random variable in its own right, rather than an unknown constant as in traditional models. As a consequence, previous experience and engineering knowledge can be included in the model by constructing an appropriate prior distribution on each parameter. A variety of vague priors have been used for deriving posteriors through Markov chain Monte Carlo simulation. Sensitivity analysis shows that these model parameters are stable over the priors assigned.

Two things should be pointed out here. Firstly, due to the absence of failure data under actual operating conditions, there is no way, to the best of the author's knowledge, to verify the stress acceleration model or the lifetime projection. Even if simulation is used, the assumptions of a lifetime distribution and acceleration are still needed. Secondly, for the same reason as just mentioned, there is no absolute rules to determine which of the Bayesian and the traditional projections is better in terms of accuracy. They represent two projections that one can choose from, depending on the amount of available sample data, the specific problem at hand, as well as one's judgment and preference. However, with the same amount of sample data, a Bayesian estimation from a non-informative prior is at least as good as a classical one. Additionally, the Bayesian approach offers the flexibility of selecting a point estimator based on the posterior to minimize the risk of a wrong decision.

c. Failure rate estimation

Chapter VII presents another way of using the Bayesian approach to reliability assessment. Breakdown data under an accelerated stress are modeled with the well-accepted bathtub failure rate curve. The DFR segment is estimated with a non-parametric Bayesian approach, while the CFR segment is estimated with a regular parametric Bayesian approach. Again, the Bayesian approach is favored because of the small sample size that is available for each estimation. The estimated DFR and CFR at various accelerated stress levels can provide a large amount of information on stress acceleration and device reliability for burn-in optimization.

To partition the sample data for the DFR estimation and the CFR estimation, an exponentiality test is performed. The test procedure includes data transformation and does not need to estimate the unknown distribution parameters. As one of the most powerful tests, it is favored because the test statistic can reveal the failure rate is increasing or decreasing if it is constant. There are other ways to determine the partition, and it would be interesting to compare them. However, the emphasis here is on the non-parametric approach for estimating a monotone failure rate. Owing to the setup of a non-parametric likelihood and a multivariate Dirichlet prior, implementation of the approach is straightforward. As less sample data are available on new devices, it is likely that a decision-maker is unable to specify or unwilling to assume the distribution function. Therefore, the potential application area of this approach is quite large.

VIII.2. Contributions

The major challenges in investigating the reliability of current and future generations of nano devices are the low failure rate and the limited amount of failure data. In this situation, the classical approach to projecting reliability under the operating conditions is inappropriate. The proposed Bayesian approach can fill this void by providing a plausible prediction for timely decision-making. Its value lies in that it can fully utilize available information in modeling uncertainty and provide cogent prediction with limited resource in a reasonable period of time. The Bayesian approach can be easily adopted to model a different device by replacing the lifetime distribution and the acceleration model with other forms of functions.

As both the time-to-market and the life cycle of a product decrease, the inclusion of reliability considerations at the design stage seems to be the best strategy for manufacturers. The Bayesian approach enables one to make early decisions based on sound judgments. It makes the best use of engineering knowledge and experts' opinions from various sources and combines them with sample data in the posterior. Furthermore, the derived posterior can be used as the prior for future studies on subsequent devices.

This study also sets up procedures of investigating the reliability of nano devices. At the beginning stages, understanding of the material properties, device physics, and failure mechanisms is indispensable to constructing valid and useful reliability models. As physics-of-failure plays a role of increasing importance, it is worthwhile to put forth an effort to study the material, electrical and reliability characteristics of the device under investigation. Engineering knowledge and the preliminary data obtained are the main sources of prior information for the Bayesian model.

The special contribution of this work is that the real failure data for high k thin films are collected from ramped and constant voltage stress tests. These data, together with the first-hand experience gained through the testing, are of great value to reliability investigation in both the physical and statistical realms. Among these data, time-to-breakdown from ALT is analyzed with Bayesian approaches in lifetime projection and failure rate estimation, whereas the I-V characteristics are used to examine dielectric relaxation and to identify breakdown mode.

Infant mortality is a serious problem that occurs on most semiconductor devices. Knowledge about the early decreasing failure rate is essential for manufacturers to establish cost-effective burn-in operations. The decreasing failure rate estimation, using the non-parametric approach presented in Chapter VII, can have an immediate impact on burn-in optimization. It can provide a fast and reliable estimation when only a small sample of data is available. In practice, this gives the manufacturer a great advantage in terms of time and cost savings.

VIII.3. Future directions

The research covered in this dissertation may be extended in the following directions.

- Asymptotically, Bayesian estimation tends to be the same as maximum likelihood estimation. However, the Bayesian approaches are most useful when sample data is rare. In this case, prior assumptions often have a large influence on the estimation. Therefore, prior distributions deserve special attention, and emphasis should be placed on how to construct a prior based on information from various sources.

- The posterior distribution of an unknown parameter allows one to choose a desired point estimator. To take advantages of this, the utility (loss) function that depicts a particular application should be defined. Then maximizing (minimizing) the utility (risk) should lead to the estimator that best suits the problem at hand. Other than symmetric functions such as the squared-error loss, asymmetric functions are probably more realistic.
- Temperature is also a key stress factor that affects the reliability of dielectric thin films. Although temperature stress is not as crucial as the electric field, it is necessary to know how dielectric films behave under different temperature levels. In reality, the electric field and temperature are interactive. Hence, reliability should be examined and modeled when both of these stress factors are considered simultaneously.

Upon written request, the author would be glad to provide the LabVIEW programs of ramped and constant voltage stress tests, and the WinBUGS models for Markov chain Monte Carlo simulation used in the dissertation.

REFERENCES

[1] E. Atanassova and A. Paskaleva, “Breakdown fields and conduction mechanisms in thin Ta_2O_5 layers on Si for high density DRAMs,” *Microelectronics Reliability,* vol. 42, pp. 157–173, 2002.

[2] B. Boukai, “Bayes sequential procedure for estimation and for determination of burn-in time in a hazard rate model with an unknown change-point,” *Sequential Analysis,* vol. 6, no. 1, pp. 37–53, 1987.

[3] B. Brar, G. D. Wilk, and A. C. Seabaugh, “Direct extraction of the electron tunneling effective mass in ultra-thin SiO_2,” *Applied Physics Letter*, vol. 69, no. 18, pp. 2728–2730, 1996.

[4] K.-M. Chang, C.-H. Li, S.-W. Wang, T.-H. Yeh, J.-Y. Yang, and T.-C. Lee, “The relaxation phenomena of positive charges in thin gate oxide during Fowler-Nordheim tunneling stress,” *IEEE Transactions on Electron Devices*, vol. 45, no. 8, pp. 1684–1689, 1998.

[5] Z. Chen, D. Ferger, and J. Mi, “Estimation of the change point of a distribution based on the number of failed test items,” *Metrika*, vol. 53, pp. 31–38, 2001.

[6] W.-T. Chien and W. Kuo, “A nonparametric Bayes approach to decide system burn-in time,” *Naval Research Logistics*, vol. 44, pp. 655–671, 1997.

[7] K. Chou and K. Tang, “Burn-in time and estimation of change-point with Weibull-exponential mixture distribution,” *Decision Sciences*, vol. 23, no. 4, pp. 973–986, 1992.

[8] R. Degraeve, G. Groeseneken, 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*, vol. 45, no. 4, pp. 904–911, 1998.

[9] R. Degraeve, B. Kaczer, and G. Groeseneken, “Degradation and breakdown in thin oxide layers: Mechanisms, models and reliability prediction,” *Microelectronics Reliability*, vol. 39, pp. 1445–1460, 1999.

[10] R. Degraeve, B. Kaczer, and G. Groeseneken, “Reliability: A possible showstopper for oxide thickness scaling?” *Semiconductor Science and Technology*, vol. 15, pp. 436–444, 2000.

[11] P. Dellaportas and A. F. M. Smith, “Bayesian inference for generalized linear and proportional hazards models via Gibbs smapling,” *Applied Statistics*, vol. 42, no. 3, pp. 443–459, 1993.

[12] D. J. DiMaria, E. Cartier, and D. Arnold, “Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon,” *Journal of Applied Physics*, vol. 73, no. 7, pp. 3367–3384, 1993.

[13] D. J. Dumin and J. R. Maddux, “Correlation of stress-induced leakage current in thin oxides with trap generation inside the oxides,” *IEEE Transactions on Electron Devices*, vol. 40, no. 5, pp. 986–993, 1993.

[14] P. Erto and M. Giorgio, “Assessing high reliability via Bayesian approach and accelerated tests,” *Reliability Engineering and System Safety*, vol. 76, pp. 301–310, 2002.

[15] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, “Effective electron mobility

in Si inversion layers in metal-oxide-semiconductor systems with a high k insulator: The role of remote phonon scattering," *Journal of Applied Physics*, vol. 90, no. 9, pp. 4587–4608, 2001.

[16] R. M. Fleming, D. V. Lang, C. D. W. Jones, M. L. Steigerwald, D. W. Murphy, G. B. Alers, Y.-H. Wong, R. B. van Dover, J. R. Kwo, and A. M. Sergent, "Defect dominated charge transport in amorphous Ta_2O_5 thin films," *Journal of Applied Physics*, vol. 88, no. 2, pp. 850–862, 2000.

[17] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–408, 1990.

[18] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Boca Raton, Florida: Chapman & Hall/CRC, 2000.

[19] G. Ghibaudo, G. Pananakakis, R. Kies, E. Vincent, and C. Papadas, "Accelerated dielectric breakdown and wear out standard testing methods and structures for reliability evaluation of thin oxides," *Microelectronics Reliability*, vol. 39, pp. 597–613, 1999.

[20] W. R. Gilks, A. Thomas, and D. J. Spiegelhalter, "Software for the Gibbs sampler," *Computational Science and Statistics*, vol. 24, pp. 438–448, 1992.

[21] G. Gouno, G. Deleuze, and M. Brizoux, "Bayesian approach of failure rate estimation in field condition through accelerated testing," in *Proceedings of 1993 Electronic Components and Technology Conference*, pp. 116–122, 1993.

[22] C. Le Gressus and G. Blaise, "Breakdown phenomena related to trapping/detrapping processes in wide band gap insulators," *IEEE Transactions on*

Electrical Insulation, vol. 27, no. 3, pp. 472–481, 1992.

[23] T. Hori, *Gate Dielectrics and MOS ULSIs: Principles, Technologies, and Applications*. Berlin: Springer-Verlag, 1997.

[24] *International Technology Roadmap for Semiconductors*, http://public.itrs.net, Semiconductor Industry Association, 2002.

[25] H. Jeffreys, *Theory of Probability*. Third edition, London: Oxford University Press, 1961.

[26] *JESD35-A: Procedure for the Wafer-level Testing of Thin Dielectrics*. JESD Solid State Technology Association, 2001.

[27] A. K. Jonscher, "Dielectric relaxation in solids," *Journal of Physics, D: Applied Physics*, vol. 32, pp. 57–70, 1999.

[28] W. Kuo, W.-T. Chien, and T. Kim, *Reliability, Yield, and Stress Burn-in*. Boston, Massachusetts: Kluwer Academic Publisher, 1998.

[29] Y. Kuo, J. Lu, J., P.-C. Liu, F. M. Daby, and J.-Y. Tewg, "Influence of the nanometer thick interface layer to electrical properties of the 10nm doped metal oxide high k dielectrics," *Proceedings of the* 2^{nd} *IEEE Conference on Nanotechnology*, pp. 251–253, 2002.

[30] Y. Kuo, J. Lu, and J.-Y. Tewg, "Tantalum nitride interface layer influence on dielectric properties of hafnium doped tantalum oxide high dielectric constant thin films," *Japanness Journal of Applied Physics*, vol. 42, no. 7A, L769–L771, 2003.

[31] Y. Kuo, H. Nominanda, M. Ristova, H. H. Lee, and J.-Y. Tewg, “Boron-doped a-Si:H thin film deposition process and applications in p-channel thin film transistor and photodiode for He-Ne laser light detection,” in *Proceedings of Thin Film Transistor Technologies VI*, PV 2002-23, pp. 292–300, the Electrochemical Society, 2002.

[32] Y. Kuo, J.-Y. Tewg, J. Donnelly, and J. Lu, ”Dopant effects on TaOx-based high k dielectric films,” in *ECS Proceedings of Semiconductor Technology Conference*, 2001-17, pp.324, 2001.

[33] B. C. Lai, N. Kung, amd J. Y. Lee, “A study on the capacitance-voltage characteristics of metal-Ta_2O_5-silicon capacitors for very large scale integration metal-oxide-semiconductor gate oxide applications,” *Journal of Applied Physics*, vol. 85, no. 8, pp. 4087–4090, 1999.

[34] P. A. W. Lewis, “Some results on tests for Possion processes,” *Biometrika*, vol. 52, no. 1 and 2, pp. 67–77, 1965.

[35] B. P. Linder, J. H., Stathis, and D. J. Frank, “Calculating the error in long term oxide reliability estimates,” in *Proceedings of the* 39th *Annual International Reliability Physics Symposium*, pp. 168–171, 2001.

[36] J. Lu, J.-Y. Tewg, Y. Kuo, and P. C. Liu, “Hafnium-Doped Tantalum Oxide High k Dielectrics,” in *Proceedings of the* 1st *International Symposium High Dielectric Constant Materials: Materials Science, Processing, Manufacturing, and Reliability Issues*, PV 2002-28, pp. 105–112, 2003.

[37] E. Lueder, “Advances in printed liquid crystal displays with plastic substrates,” in *Proceedings of Thin Film Transistor Technologies V*, PV 2000-31, pp. 1–10, the Electrochemical Society, 2000.

[38] W. Luo, Y. Kuo, and W. Kuo, “Dielectric relaxation and breakdown detection of doped tantalum oxide high k thin films,” *IEEE Transactions on Device and Materials Reliability*, accepted in August 2004.

[39] T. P. Ma, “Scaling of gate dielectrics for advanced CMOS applications,” *The Physics and Chemistry of* SiO_2 *and the Si-*SiO_2 *Interface - 4*, vol. 2, pp. 19–32, 2000.

[40] A. Martin, P. O’Sullivan, and A. Mathewson, “Dielectric reliability measurement methods: A review,” *Microelectronics Reliability*, vol. 38, no. 1, pp. 37–72, 1998.

[41] H. F. Martz and R. A. Waller, *Bayesian Reliability Analysis*. New York: John Wiley & Sons, 1982.

[42] National Research Council, *Implications of Emerging Micro- and Nanotechnologies*, Washington, D. C.: The National Academies Press, 2002.

[43] T. A. Mazzuchi and N. D. Singpurwalla, “A Bayesian approach to inference for monotone failure rates,” *Statistics and Probability Letters*, vol. 37, pp. 135–141, 1985.

[44] R. M. Neal, “Slice sampling,” *Technical Report*, no.2005, Department of Statistics, University of Toronto, Canada, 2000.

[45] W. Nelson, *Accelerated Testing*. New York: John Wiley & Sons, 1990.

[46] T. Nigam, R. Degraeve, M. M. Heyns, and H. E. Maes, “Constant current charge-to-breakdown: Still a valid tool to study the reliability of MOS structures?” in *Proceedings of the* 36th *Annual International Reliability Physics Symposium*, pp. 62–69, 1998.

[47] K. H. Oh, C. L. Gressus, H. Gong, C. K. Ong, G. B. T. Tan, and X. Z Ding, "Dielectric polarization relaxation measurement in SiO_2 by means of a scanning electron microscope technique," *Journal of Applied Physics*, vol. 74, no. 2, pp. 1250–1255, 1993.

[48] M. Ohring, *Reliability and Failure of Electronic Materials and Devices*. San Diego, California: Academic Press, 1998.

[49] P. K. Pathak, A. K. Singh, and W. J. Zimmer, "Bayes estimation of hazard & acceleration in accelerated testing," *IEEE Transactions on Reliability*, vol. 40, no. 5, pp. 615–621, 1991.

[50] H. Reisinger, G. Steinlesberger, S. Jakschik, M. Gutsche, T. Hecht, M. Lenonhard, U. Schröder, H. Seidl, and D. Schumann, "A comparative study of dielectric relaxation losses in alternative dielectrics," *2001 IEDM Technical Digest*, pp. 267–270, International Electron Devices Meeting, 2001.

[51] J. L. Romeu, "Anderson-Darling: A goodness-of-fit test for small samples assumptions" *Selected Topics in Assurance Related Technologies*, vol. 10, no. 5, http://www.rac.alionscience.com/pdf/A_DTest.pdf, Reliability Analysis Center, 2004.

[52] M. Schumacher, R. Waser, "Curie-von Schweidler behavior observed in ferroelectric thin films and comparison to superparaelectric thin film materials," *Integrated Ferroelectrics*, vol. 22, pp. 109–120, 1998.

[53] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, *WinBUGS User Manual*, Version 1.4, http://www.mrc-bsu.cam.ac.uk/bugs, 2003.

[54] J. H., Stathis, "Physical and predictive models of ultra thin oxide reliability in

CMOS devices and circuits," in *Proceedings of the* 39^{th} *Annual International Reliability Physics Symposium*, pp. 132–149, 2001.

[55] M. A. Stephens, "Tests based on EDF statistics," in *Goodness-of-Fit Techniques*. R. B. D'Agostino and M. A. Stephens, Ed., New York: Marcel-Dekker, 1986.

[56] M. A. Stephens, "Tests based on the exponential distribution," in *Goodness-of-Fit Techniques*. R. B. D'Agostino and M. A. Stephens, Ed., New York: Marcel-Dekker, 1986.

[57] J. S Suehle, "Reliability characterization of ultra-thin film dielectrics," in *Characterization and Metrology for ULSI Technology: 1998 International Conference*, pp. 115–119, 1998.

[58] S. M. Sze, *Semiconductor Devices: Physics and Technology*. Second edition, New York: John Wiley & Sons, 2002.

[59] P. A. Tobias, *Applied Reliability*. New York: Van Nostrand Reinhand, 1986.

[60] W. D. Westwood, N. Waterhouse, and P. S. Wilcox., *Tantalum Thin Films*. New York: Academic Press, 1975.

[61] G. D. Wilk, R. M. Wallace, and J. M. Anthony, "High k gate dielectrics: Current status and materials properties considerations," *Journal of Applied Physics*, vol. 89, no. 10, pp. 5243–5275, 2001.

[62] E. Y. Wu, W. W. Abadeer, L.-K. Lai, S.-H. Lo, and G. R. Hueckel, "Challenges for accurate reliability projections in the ultra-thin oxide regime," in *Proceedings of the* 37^{th} *Annual International Reliability Physics Symposium*, pp. 57–65, 1999.

[63] E. Y. Wu, J. H. Stathis, and L.-K. Han, "Ultra-thin oxide reliability for ULSI applications," *Semiconductor Science and Technology*, vol. 15, pp. 425–435, 2000.

VITA

Wen Luo received a B.E. in precision instruments from Tsinghua University, Beijing, China, and a M.S. in industrial engineering from the University of Washington, Seattle. She received a Ph.D. from the Department of Industrial Engineering at Texas A&M University. Her research field is the reliability of nano- and micro-electronics. Her interests are in statistical methods of reliability modeling and prediction, as well as reliability testing and characterization of nano devices. Wen Luo can be reached at 406 Ferris Hall, Department of Electrical and Computer Engineering, University of Tennessee, Knoxville, Tennessee 37996.

This document was typeset in LaTeX by Wen Luo.