

ABSTRACT

Title of dissertation: A NEW PHYSICS-OF-FAILURE BASED VLSI
CIRCUITS RELIABILITY SIMULATION
AND PREDICTION METHODOLOGY

JIN QIN, Doctor of Philosophy, 2007

Dissertation directed by: Professor Joseph B Bernstein
Department of Mechanical Engineering

It has long been a challenge for reliability engineers to provide accurate VLSI circuits reliability simulation and prediction. The decreasing feature sizes, coupled with non-ideal voltage scaling, raises new reliability concerns such as negative bias temperature instability (NBTI) and adversely affects those long-existed failure mechanisms: electromigration (EM), hot carrier degradation (HCD) and time dependent dielectric breakdown (TDDB). The multiple failure mechanisms effect, together with the increasing circuit complexity make the prediction more difficult to tackle with.

A new physics-of-failure based VLSI circuit reliability prediction methodology is proposed to handle the simulation and prediction challenges. The new methodology takes an unique top-down, bottom-up approach to reduce the modeling and simulation complexity. Detailed application breakdown reveals the cell's operation profile. Cell-level reliability characterization provides accurate operation-based dynamic stress modeling by utilizing the physics-of-failure models. For each failure

mechanism, the best-fit lifetime distribution is selected to provide reliability prediction. The application-specific circuit reliability is further predicted by considering the system structure.

A 90nm 64Kb SRAM module is designed and used as an example to demonstrate the prediction methodology. With the given application profile, simulation results showed that TDDB is the most serious reliability concern for the SRAM bit cell, NBTI is in the second place, and HCD has a negligible degradation effect. The memory core's reliability prediction shows the core has a low constant failure rate ($2.90E-4$ FIT) before $5.8E+4$ hours, and an increasing failure rate after that because NBTI wearout starts to kick in.

A NEW PHYSICS-OF-FAILURE BASED VLSI CIRCUITS
RELIABILITY SIMULATION AND PREDICTION
METHODOLOGY

by

JIN QIN

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee:

Professor Joseph B Bernstein, Chair/Advisor
Professor Martin Peckerar
Professor Carol Smidts
Professor Peter Sandborn
Professor Patrick McCluskey

© Copyright by
JIN QIN
2007

Dedication

To my beautiful wife Mingqin.

ACKNOWLEDGMENTS

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I'd like to thank my advisor, Professor Joseph Bernstein for giving me an invaluable opportunity to work on this challenging and extremely interesting project over the past five years. He has always made himself available for help and his brilliant ideas always inspired me. It has been a great pleasure to work with and learn from such an extraordinary individual.

I would like to give my special thanks to Dr. Peckerar, who provided so many insightful suggestions and gave me the permission to access the design kit. He was generous to share his knowledge on any semiconductor question I asked.

I would also like to thank Dr. Smidts, Dr. Sandborn and Dr. McCluskey, for serving on my dissertation committee. Their profound knowledge and enlightening discussions instilled many essential ingredients into this dissertation work. Their persistent support and invaluable guidance proved to be indispensable improving my research.

Furthermore, I would like to thank the other talented researchers in our group, which include Xiaojun Li, Aydin Balkan, Bo Yang, Wei Wu, Moshe Gurfinkel, Bing Huang and Xiaohu Zhang, for their help and support to this work.

Table of Contents

List of Tables	vii
List of Figures	viii
List of Abbreviations	x
1 Introduction	1
1.1 Reliability Challenges	2
1.1.1 Device Level Concerns	2
1.1.2 System Level Concerns	6
1.1.2.1 Conventional Product Qualification	6
1.1.2.2 Conventional Reliability Prediction	8
1.1.2.3 Product Reliability Trend	9
1.2 Dissertation Overview	11
2 Time Dependent Dielectric Breakdown	13
2.1 Introduction	13
2.2 Physics of Failure	14
2.2.1 Defects Inside SiO_2 Film and At The Silicon-Oxide Interface	14
2.2.2 Tunneling Currents	16
2.2.2.1 Fowler-Nordheim Tunneling	16
2.2.2.2 Direct Tunneling	17
2.2.2.3 Trap Assisted Tunneling and Other Effects	18
2.2.3 Trap Generation Mechanisms	18
2.2.3.1 Thermochemical Model	19
2.2.3.2 AHI Model	21
2.2.3.3 AHR Model	23
2.2.3.4 Voltage Dependence of Defect Generation	25
2.3 TDDDB Models	25
2.3.1 Lifetime Models	25
2.3.1.1 Voltage/Field Dependence	25
2.3.1.2 Temperature Dependence	26
2.3.1.3 Area Dependence	28
2.3.2 Statistical Models	28
2.3.2.1 Percolation Theory	28
2.3.2.2 Weibull Distribution	30
3 Negative Bias Temperature Instability	33
3.1 Introduction	33
3.2 Physics of Failure	34
3.2.1 Interface Trap Generation: Reaction-Diffusion Model	35
3.2.2 Fixed Charge Generation	37
3.2.3 Recovery and Saturation	38

3.3	NBTI Models	39
3.3.1	Degradation Models	39
3.3.2	Lifetime Models	40
3.3.3	Statistical Model	41
4	Hot Carrier Degradation	42
4.1	Introduction	42
4.2	Physics of Failure	42
4.2.1	Channel Hot Electron (CHE) Injection	44
4.2.2	Drain Avalanche Hot Carrier (DAHC) Injection	45
4.2.3	Secondary Generated Hot Electron (SGHE) Injection	47
4.3	HCD Models	48
4.3.1	Lucky Electron Model (LEM)	48
4.3.2	Empirical Lifetime Models	52
4.3.2.1	Power Law Model	52
4.3.3	Dynamic Stress Lifetime Modeling	54
4.3.4	Temperature Dependence Model	55
4.3.5	Statistical Model	56
5	Electromigration	57
5.1	Introduction	57
5.2	Physics of Failure	58
5.3	EM Models	65
5.3.1	Lifetime Models	65
5.3.1.1	Black's Equation	65
5.3.1.2	EM Lifetime Dependence On Interconnect Width and Length	67
5.3.1.3	EM Lifetime Dependence On Current Waveforms	69
5.3.1.4	EM Lifetime Dependence On Joule Heating Effect	70
5.3.2	Statistical Models	71
5.3.2.1	Lognormal Distribution	71
5.3.2.2	Weibull Distribution	74
5.3.2.3	Bimodal Distribution	74
6	Physics-of-Failure Based VLSI Circuits Reliability Prediction Methodology	77
6.1	Introduction	77
6.2	Review of Reliability Prediction Tools	79
6.3	Methodology	81
6.3.1	Assumptions	86
6.3.2	Input Data	87
6.3.3	Device Thermal Analysis	89
6.3.4	Cell Reliability Estimation	91
6.3.4.1	Dynamic Stress Profile Extraction	91
6.3.4.2	Equivalent-Stress-Factor Evaluation	92
6.3.4.3	Cell Reliability	94

6.3.5	Chip Reliability Prediction	96
6.3.5.1	Functional Block Reliability	97
6.3.5.2	Power Network EM Estimation	97
6.4	Case Study: SRAM Reliability Prediction	98
6.4.1	SRAM Design	98
6.4.2	Reliability Prediction	102
6.4.2.1	Application Profile	102
6.4.2.2	Chip Thermal Analysis	102
6.4.2.3	SRAM Cell Reliability Characterization	102
6.4.2.4	Power Bus Reliability Estimation	104
6.4.2.5	Memory Core Reliability Prediction	104
6.5	Summary	107
7	Summary	108
7.1	Results	108
7.2	Future Work	110
7.2.1	Model Verification and Validation	110
7.2.2	Prediction Tool Development	111
A	Sample Scripts and Programs	112
A.1	Dynamic Stress Profile Extraction Script	112
A.2	Example: MATLAB Script for Device Reliability Estimation	114
	Bibliography	120

List of Tables

2.1	Reported values of the power law exponent n	27
6.1	SRAM cell operation profile	102
6.2	TDDDB, NBTI and HCD lifetime models	103
6.3	ESFs of SRAM cell transistors	103
6.4	Cell transistor stress duty factors	104

List of Figures

1.1	Failure rates of systems from two avionics suppliers (90% confidence interval). Systems are grouped by type and year-into-service.	10
1.2	Manufacturers' reliability data. Data sources: [1], [2] and [3]	11
2.1	Location and identification of charges in SiO_2 and at the oxide-silicon surface [4].	15
2.2	Schematic illustration of the Fowler-Nordheim tunneling current and direct tunneling current [5].	17
2.3	Outline of the mechanism of defect generation leading to breakdown in SiO_2 [6].	19
2.4	Band diagram models of AHI model [4].	22
2.5	Outline of neutral electron trap generation [5].	24
2.6	The percolation model for oxide breakdown [5].	29
2.7	Weibull shape parameters vs. oxide thickness [7].	31
3.1	Schematic two-dimensional representation of the $S_i - SiO_2$ interface, showing (a) the $\equiv S_i$ defect, (b) how this defect may be electrically activated during NBTI to form an interface trap, a fixed oxide charge, and a hydroxyl group, and (c) the OH diffuses through the oxide. Adapted from [8].	36
4.1	Hot carrier generation and degradation in MOSFETs. Adapted from [9].	44
4.2	Channel-hot-electron injection. Occurs when $V_g \approx V_d$	46
4.3	Drain avalanche hot-carrier injection. Occurs when $V_d > V_g$	46
4.4	Secondarily generated hot electron injection.	47
5.1	Summary of EM failure process [10].	59
5.2	Schematic of grain-boundary triple point configuration [4].	64

6.1	Flowchart of the physics-of-failure based statistical reliability prediction method.	85
6.2	Schematic of one-bit 6T SRAM cell. Store/Storen represent cell state. WORD line enables pass transistors M5 and M6 during memory read and rite.	99
6.3	Block diagram of the simulation circuit in SRAM bit cell reliability characterization. The circuit consists of one 6T cell, read/write control logic and output sense amplifier.	100
6.4	SRAM cell SPICE simulation stimuli. PRE is the precharge which exerts before each read/write operation. WORD signal controls cell transistor M5 and M6 during read and write. DATA carries the bit info “0” or “1” during each write operation.	101
6.5	SRAM core reliability prediction.	106
6.6	SRAM core failure rate prediction.	106

List of Abbreviations

ALT	Accelerated Lifetime Test
AVSI	Aerospace Vehicle Systems Institute
DAHC	Drain Avalanche Hot Carrier
CFR	Constant Failure Rate
CHE	Channel Hot Electron
EM	Electromigration
ETM	Effective Temperature Model
ESS	Environment Stress Screen
ESF	Equivalent Stress Factor
FN	Fowler-Nordheim
HCD	Hot Carrier Degradation
HTOL	High Temperature Operation Life
IC	Integrated Circuit
LEM	Lucky Electron Model
MTTF	Mean Time to Failure
MLE	Maximum Likelihood Estimation
NBTI	Negative Bias Temperature Instability
PoF	Physics of Failure
RF	Radio Frequency
RC	Resistance-Capacitance
TDDB	Time Dependent Dielectric Breakdown

Chapter 1

Introduction

The semiconductor industry has witnessed big technology achievements and exceptional market growth in the last 40 years. Nowadays, microelectronic devices that contain millions of transistors can be easily found in numerous appliances like gps receivers, cellphones, mp3 players and even rice-cookers. The maturing of the semiconductor market has shifted the industry's emphasis from previous defense and aerospace markets that require high reliability performance to commercial and consumer markets in which timing and functionality are now the priorities. Marketing pressure and fierce competition drive the manufacturers to keep introducing new materials, new processes and novel devices. Many aspects of semiconductor design and manufacturing are undergoing dramatic changes that may threaten the high level of reliability that customers experienced in the past. Designers are under extensive pressure to have their designs work the first time with acceptable reliability. They need to tailor the tradeoff between performance and lifetime to meet the needs of different market segments. For manufacturers, product reliability assurance is facing big challenges since the device complexity is increasing and the gap between normal operating and accelerated test conditions is narrowing. To handle these challenges, designers and manufacturers must be equipped with accurate reliability simulation and prediction tools in order to achieve the reliability performance

goal from both circuit level and system level.

1.1 Reliability Challenges

1.1.1 Device Level Concerns

Advances in CMOS semiconductor technology have been driven by aggressive device scaling to achieve density, speed, and power improvement. High-resolution lithography and ion implantation are two important techniques in device scaling [11]. Under ideal scaling in which device voltages and device dimension (both horizontal and vertical) are scaled by the same factor ($k = 0.7$), gate delay decreases by 30% from one generation to the next, transistor density doubles, and dynamic power per transistor decreases by about 50%. However, supply voltages are not scaled ideally because of the subthreshold non-scaling [11] and reluctance to depart from the standardized voltage levels of the previous generation in order to retain competitive frequency growth. The non-ideal voltage scaling together with the reduced feature sizes creates higher electric fields in oxide and channel, higher current densities and higher power density. The high power density brings high temperatures that will adversely affect most of the intrinsic failure mechanisms that exhibit exponential or larger dependence on temperature. Simulation had shown that temperature of the hottest structure on a POWER4-like processor increases by an average 15 degrees Kelvin from 180nm to 65nm [12]. Besides the high temperature, the high electric fields cause big threats to reliability.

Gate Dielectric

Silicon dioxide (SiO_2), with a bulk resistivity of $\sim 10^{15} \Omega\text{-cm}$ and a dielectric breakdown strength of $\sim 10^7 \text{ V/cm}$ [13], is the key to the success of the semiconductor industry due to its process advantage and high quality. Scaling has driven t_{ox} (thickness of SiO_2) down to the fundamental limiting point $\sim 1.2 \text{ nm}$, since the wave properties of the electron mean that oxides thinner than 1 nm are no longer insulated and oxides much thinner than 1 nm may never find extensive use in ICs because of excessively high oxide currents [14]. For SiO_2 , the reliability threat comes from the defect generation whose rate is proportional to the leakage current density flowing through the oxides. The leakage current, either Fowler-Nordheim (FN) tunneling or direct tunneling, is controlled by quantum-mechanical tunneling. Higher fields increase the leakage current and accelerate dielectric degradation and eventually cause breakdown. Stathis and DiMaria [15] showed that the direct tunneling current increases exponentially by about one order of magnitude for every $0.2\text{--}0.3\text{-nm}$ reduction in oxide thickness for $t_{ox} \leq 4 \text{ nm}$. This means further scaling of t_{ox} will reduce the oxide lifetime by $10^{\frac{\Delta t_{ox}}{0.22}}$, where Δt_{ox} is the reduction of t_{ox} in nanometers [12]. Furthermore, the Weibull slope β of dielectric lifetime distribution was realized to become smaller as t_{ox} decreases and approach 1 for $t_{ox} \leq 2 \text{ nm}$ [16, 17, 18]. The Weibull distribution now reduces to an exponential distribution that has a constant failure rate. This means burn-in will not be effective in reducing the intrinsic failure rate of dielectric breakdown during operation since it is a purely random phenomena. Failures will happen during

the normal operation for ultrathin oxides while most of the failures occur near the end of life for thick oxides with larger β . The above two factors will cause the increase of failure rates of chips manufactured by advanced technologies with ultrathin oxides.

Adding nitrogen to oxide is a temporary solution for achieving reasonable physical thickness and acceptable reliability performance. High-k dielectric will eventually be needed once its breakdown characteristics have been figured out. These technology improvement options are still under investigation and may introduce new reliability concerns.

Negative Bias Temperature Instability

Negative bias temperature instability (NBTI) has been known since the early days of MOS device development [19]. Compared with hot carrier degradation (HCD), NBTI remained marginal for many years until the migration to nitrided oxide and surface-channel devices [20]. Even though the exact root causes of NBTI are not yet well understood, it is commonly admitted that under a negative gate voltage and an elevated temperature a build-up of positive charges occurs at the $Si - SiO_2$ interface and in the oxide layer leading to the reduction of PMOSFET performances, like a shift in the threshold voltage (V_{th}) and a decrease in mobility. The interface trap density induced by NBTI increases with decreasing oxide thickness [21]. This t_{ox}^{-1} dependence of the interface-trap, together with the increasing oxide fields, operating temperature and decreasing voltage headroom ($V_g - V_{th}$, V_g is the gate voltage, V_{th}

is the threshold voltage), implies NBTI becomes more severe for ultrathin oxides. And more importantly, the inverse condition in which PMOSFET suffers NBTI mostly arises universally in inverting logic, SRAM cells, I/O system and dynamic logic. NBTI has become the pervasive concern of the semiconductor industry in recent years.

Hot Carrier Degradation

For a short channel device with non-ideally scaled voltages, hot carrier degradation (HCD) tends to worsen as the channel and gate fields increase following the scaling. Transistors will suffer more drain avalanche hot carrier (DAHC) injection as the operating frequency increases. The system-on-chip (SOC) integration which contains digital, analog, mixed-signal and even radio frequency (RF) functions, may introduce more hot carrier vulnerable parts into the chip. And, the reach of minimal voltage levels makes HCD an increasing concern.

Interconnect

To meet the need of continually reducing the resistance-capacitance (RC) time delays, copper interconnect and low-k interlevel dielectrics gradually replace aluminum interconnect and SiO_2 dielectrics. The introduction of copper interconnect involves new process complications and interactions, such as dual damascene patterning, barriers to contain the copper, copper vias, high via aspect ratios and complex interaction with low-k interlevel dielectrics [22]. In addition to these process changes, the interconnect density, number of layers,

power consumption, and self-heating increased. These factors bring more reliability issues that must be dealt with in scaling. The most critical concerns include at least the following: stress migration of copper vias and lines, TDDB of the low k dielectric, copper and via electromigration (EM) performance and issues due to the porous nature of the low-k dielectrics and moisture ingress. Copper EM generally occurs along the copper/capping layer interface. For the same current density, EM performance is expected to degrade with scaling because of the relative increase in copper-interface area vs. volume.

1.1.2 System Level Concerns

System reliability becomes more difficult to tackle since there are so many challenges at the device level as stated above. Burn-in, highly accelerated lifetime testing (HALT) and high temperature operation life (HTOL) are extensively utilized by semiconductor manufacturers for screening out outliers and monitoring the product reliability. The increasing device density and power dissipation, more severe degradation for scaled devices, together with limited lifetime testing resources, make reliability qualification more difficult than ever before.

1.1.2.1 Conventional Product Qualification

To guarantee the reliability performance in the field, semiconductor manufacturers take multiple approaches to screen and test products before shipment. Those approaches include burn-in tests, HTOL tests, environment stress screens (ESS)

and other kinds of accelerated lifetime tests (ALT). These methods were developed for long-channel devices that have the failure characteristics described by the bathtub curve. Some of these methods are not as effective as before for the advanced technologies. The acceleration model behind these traditional ALT is

$$AF_S = AF_T \cdot AF_V \quad (1.1)$$

where AF_S is the system acceleration factor, AF_T is the temperature acceleration factor and AF_V is the voltage acceleration factor. Generally, manufacturers only obtain one activation energy and one voltage acceleration parameter from high temperature, high voltage tests at each technology node. These parameters are then used in qualification test planing and reliability extrapolation. This kind of acceleration modeling stands only where one failure mechanism dominates at both the normal usage and accelerated conditions or all failure mechanisms have the same temperature and voltage acceleration model and parameters – generally this is not true. For deep submicron devices, all the intrinsic failure mechanisms will cause failures but they don't have the same temperature and voltage acceleration parameters. In this multiple failure mechanisms era, AF_S can't be modeled accurately with only one set of temperature and voltage acceleration parameters because of those parameters' stress-dependence [23]. Using estimated parameters from high temperature, high voltage acceleration tests to extrapolate failure rate at normal use conditions will get a better-than-real result.

For ultrathin gate oxides, the ALT methods become less effective because dielectric breakdown is more of a random phenomenon rather than a wearout failure.

The high voltage, high temperature stresses that was applied to screen out oxides with detrimental defects may introduce damage to the ‘health’ oxides and cause an increase of the failure rate.

Further, today’s qualification tests are designed to meet the requirements of commercial and consumer markets in which products have short life cycles of 3-5 years. The long lifetime and low operation failure rate that are required for the defense and aerospace industries can’t be guaranteed based on these tests. Avionics designers need a comprehensive chip-level reliability prediction tool to help them design in reliability.

1.1.2.2 Conventional Reliability Prediction

The well-known military handbook for reliability prediction of electronic equipment, MIL-HDBK-217F [24], was an empirical method based on the constant failure rate (CFR) model [25]. Although the CFR model was used without physical justification, it is not difficult to reconstruct the rationale for the use of the CFR model that mathematically describes the failure distribution of systems in which failures are due to completely random or chance events [26]. This concept was adapted by many other reliability prediction methods at that time, like Bellcore RPP, British Telecom HRD4, NTT Procedure and Siemens Procedure [27]. There are some disadvantages of those handbook-based prediction methods: (1) lack of connection with the physics-of-failure (PoF) approach; (2) inability to model the infant mortality or wearout ; (3) lagging far behind the technology advancements.

Nowadays, circuit designers have reliability simulators as an integral part of the design tools, like Cadence Ultrasim [28] and Mentor Graphics Eldo [29]. These simulator models the most significant physical failure mechanisms and help designers meet the lifetime performance requirement. However, there are disadvantages which hinder designers' adoption of these tools. First, these tools are not fully integrated into the design software because the full integration requires technical supports from both the tool developers and the foundry. Second, these tools can't handle the large-scale design efficiently. Modern chips are composed of tens or hundreds of millions of transistors. The increasing complexity makes it impossible or prohibitively expensive to exercise full scale simulation considering the resource that simulation will consume. Chip level reliability prediction tools, today, only focus on the chip's end-of-life (EOL), when the known wearout mechanisms will kick-in and dominate. However, these prediction tools do not predict the random, post burn-in failure rate that would be seen in the field.

1.1.2.3 Product Reliability Trend

For scaled semiconductor devices, the possible reliability threats have been briefly discussed. Besides these theoretical analysis, field data should be examined to disclose the real trend. The two sets of data below give some hints of how scaling adversely affects device reliability.

The first set of data is avionics field failure data which was analyzed by the Center of Reliable Electronic Systems (CRES) of the University of Maryland, Col-

lege Park [30]. The original failure records were provided by two members of the Aerospace Vehicle Systems Institute (AVSI) Project #17 team. Based on the assumption that the system with a latter service entry used newer devices with smaller geometries, the failure records were compiled, categorized and statistically analyzed according to the system type and year-into-service. Likelihood ratio tests showed that exponential distribution fits most of the data subset and all the failure rates are shown in Fig. 1.1. The result shows that systems D, E, F and G show an increasing trend after 1994. System A shows the same trend after 1998, system H is the outlier, which demonstrated a decreasing failure rate trend.

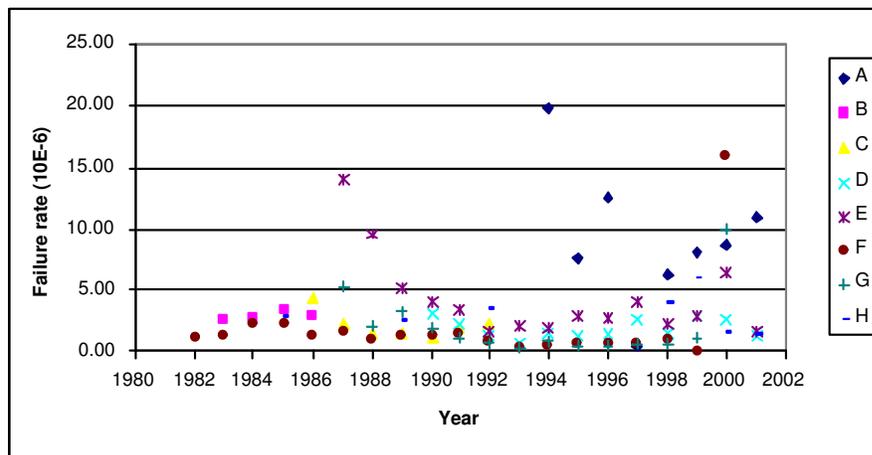


Figure 1.1: Failure rates of systems from two avionics suppliers (90% confidence interval). Systems are grouped by type and year-into-service.

The second set of data has three sources: (1) manufacturers' acceleration test data collected from literatures [1], [2]; (2) acceleration test which was coordinated by CRES and (3) burn-in data of one major semiconductor manufacturer. Device failure rates were normalized to normal use conditions by extrapolation with appro-

appropriate acceleration models and parameters (from literature or manufacturer). As shown in Fig. 1.2, device wearout is clearly observed and also approaching fast as technology advanced from 180nm to 90/65nm.

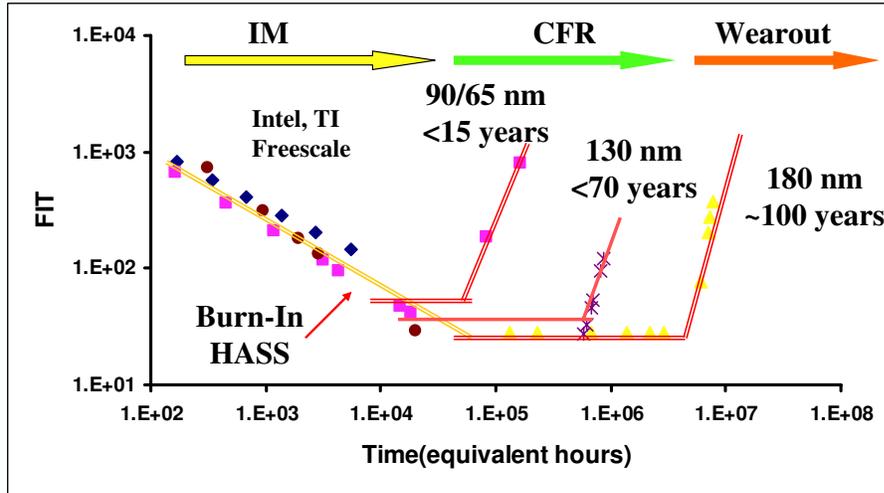


Figure 1.2: Manufacturers' reliability data. Data sources: [1], [2] and [3]

1.2 Dissertation Overview

To deal with these reliability challenges at the circuit and system level, accurate reliability models and tools for lifetime estimation should integrate the physics-of-failure approach and the statistical approach. The failure mechanism driven PoF approach must be employed to identify the potential failures and evaluate their kinetics and impact based on the specific application conditions and market requirements. The statistical approach is especially necessary for product/system engineers to model the variation, design accelerated qualification plans and make reliability predictions from individual failure mechanisms to a comprehensive product reliability.

The purpose of this dissertation is to develop a PoF based VLSI circuit reliability simulation and prediction methodology that combines the advantages of the PoF approach and the statistical approach. After this introduction, the next four chapters (Ch. 2, 3, 4 and 5) discuss the four intrinsic failure mechanisms: TDDB, NBTI, HCD and EM, respectively. The physics of failure behind these failure mechanisms, physical and statistical models will be covered in detail. Chapter 6 introduces the PoF based reliability simulation and prediction method. Chapter 7 summarizes the results and discusses future work.

Chapter 2

Time Dependent Dielectric Breakdown

2.1 Introduction

Time Dependent Dielectric Breakdown (TDDB) is a wear-out phenomenon of silicon dioxide, the thin insulating layer between the control “gate” and the conducting “channel” of the transistor. SiO_2 has a very high bandgap of about 9 eV and an excellent scaling and process integration capability that makes it the key factor of the success in MOS-technology. Layers as thin as 1.5 nm can be obtained and implemented as gate dielectrics in fully functioning MOSFETs with gate lengths of only 40 nm [31]. Although SiO_2 has many extraordinary properties, it is not perfect and suffers degradations which are caused by stress factors such as voltage and high oxide field. TDDB, one of the final results of oxide degradation, has been the subject of numerous studies that were published over the past four decades. But even today, a complete understanding of TDDB has not yet been acquired. Basic models, such as **E**-model, **1/E**-model and power law model, have been proposed to model the lifetime. Weibull distribution has been shown to be the best lifetime distribution since TDDB appears to be a “weakest link” type of failure mechanism. The percolation theory has been successfully applied to the statistical description of TDDB.

2.2 Physics of Failure

TDDDB mechanisms are still under extensive investigation because of the scaling down of oxide thickness and the introduction of new materials. The general idea is that a driving force such as the applied voltage or the resulting tunneling electrons create defects in the volume of the oxide film [32]. The defects accumulate with time and eventually reach a critical density triggering a sudden loss of dielectric properties. A surge of current produces a large localized rise in temperature leading to permanent structural damage in the silicon oxide film.

2.2.1 Defects Inside SiO_2 Film and At The Silicon-Oxide Interface

The silicon dioxide is not ideally perfect as there are charges inside the oxide and near the oxide-silicon interface. These charges can be mobile ionic charges, electrons, or holes trapped in the oxide layer. They also can be fabrication-process-induced fixed oxide charges near the oxide-silicon interface, and charges trapped at the surface states at the oxide-silicon interface [4]. Electrons and holes can make transitions between the crystalline states near the oxide-silicon interface to the surface states. These charges will definitely affect the electrical characteristics of devices and are an important factor in the TDDDB. Fig. 2.1 shows the names and locations of charges inside silicon dioxide and at the silicon-oxide interface.

1. Interfacial oxide charge. It is located within 0.2 nm of the $SiO_2 - Si$ surface.

The interfacial oxide charge arises from oxidation-induced structural defects, metal impurities, and broken bonds due to charge injection. These interfacial

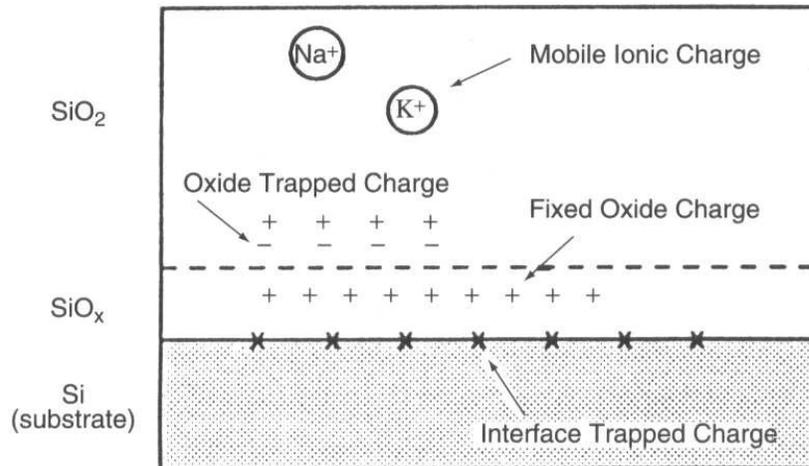


Figure 2.1: Location and identification of charges in SiO_2 and at the oxide-silicon surface [4].

states are amphoteric; they are acceptor-like in the upper half of the Si band gap, and donor-like in the lower half of the band gap.

2. Fixed oxide charge. The fixed oxide charge is a positive charge located some 3 to 5 nm from the Si-SiO_2 interface. It is primarily due to excess silicon species introduced during oxidation and postoxidation heat treatment. It is fixed and largely uninfluenced by the normal operating voltages of the MOS transistor.
3. Oxide trapped charge. This charge is distributed throughout the bulk of the SiO_2 film. Sources of this include the oxide growth process, fabrication of device [13] and high-energy electrons. The fabrication introduced charge can be removed through low-temperature annealing.
4. Mobile Na^+ and K^+ ionic charges. These charges have been virtually elimi-

nated as a source of a reliability problem.

It is the dynamic generation and the alternation of existing oxide charge states under high electric fields that ultimately cause dielectrics to breakdown. There are several driving forces which involve the charge generation and alternation: Fowler-Nordheim tunneling, direct tunneling, and trap assisted tunneling.

2.2.2 Tunneling Currents

2.2.2.1 Fowler-Nordheim Tunneling

Fowler-Nordheim tunneling is a quantum mechanical tunneling process where the electrons can penetrate through the oxide barrier into the conduction band of the oxide. The complete theory of Fowler-Nordheim tunneling is rather complicated and not discussed here. Fig. 2.2 (a) illustrates electron tunneling from the silicon surface inversion layer to the SiO_2 conduction band. The Fowler-Nordheim tunneling current density is given by

$$J_{FN} = \frac{q^3 \varepsilon_{ox}^2}{16\pi^2 \hbar \phi_{ox}} \exp\left(-\frac{4\sqrt{2m^*} \phi_{ox}^{\frac{3}{2}}}{3\hbar q \varepsilon_{ox}}\right), \quad (2.1)$$

where ε_{ox} is the electric field in the oxide, ϕ_{ox} is the silicon-silicon dioxide interface potential barrier for electrons, m^* is the electron effective mass, \hbar is Planck's constant. Eq. 2.1 shows that the Fowler-Nordheim tunneling current is characterized by a straight line in a plot of $\log(J/\varepsilon_{ox}^2)$ versus $1/\varepsilon_{ox}$.

The Fowler-Nordheim tunneling current is dependent on the oxide field, thus the voltage applied to the gate oxide. It can occur in most any gate oxide, provided

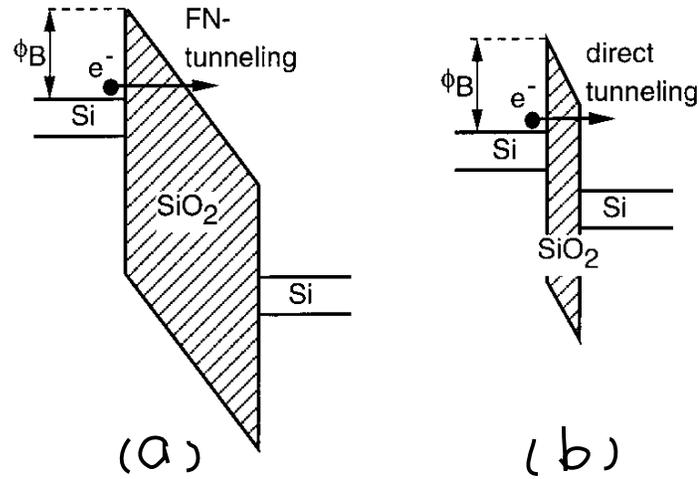


Figure 2.2: Schematic illustration of the Fowler-Nordheim tunneling current and direct tunneling current [5].

the voltage is sufficient for electrons to tunnel through the barrier.

2.2.2.2 Direct Tunneling

Direct tunneling is the dominant current conduction mechanism through sub-3nm oxide layers [33]. It is also a quantum mechanical process and much more complicated than the Fowler-Nordheim tunneling. When the oxide voltage drops below 3 V, the electrons tunnel directly from the anode to the cathode as shown in Fig. 2.2 (b). There is no simple dependence of the tunneling current density on voltage or electric field and no easy closed analytic form of expression. The direct tunneling current can be very large for thin oxide layers as shown by Lo [34]. A detailed review of direct tunneling current models can be found in [35].

2.2.2.3 Trap Assisted Tunneling and Other Effects

Trap assisted tunneling occurs when electrons tunnel through the oxide into traps (empty bonding sites on silicon dioxide molecules) and then from the traps into the silicon. Trap assisted tunneling depends on the density of the traps and the electric field and is rather complex to formalize.

Other factors can influence the behavior of the tunneling current. One such effect occurs due to the gate-drain overlap region and the gate-source overlap region. Engineers must take valence band tunneling into account. Valence band tunneling becomes more important with very thin oxides. Another mechanism that has been observed is electron hopping, caused by the jump of thermally excited electrons between isolated states. Field emission, or the tunneling of trapped electrons to the conduction band, is another factor affecting the tunneling current. Finally, an effect called Poole-Frenkel emission, the tunneling of trapped electrons into the conduction band due to barrier lowering, can affect the overall gate current.

2.2.3 Trap Generation Mechanisms

Trap generation is the key factor determining the oxide degradation and breakdown. Electrons flowing across the oxide trigger several processes depending on their energy, which is determined by the gate voltage for thin oxides where electron transport is ballistic or quasi-ballistic. The generation and buildup of microscopic defects that act as electron traps cause oxide failures when the sufficient defect density is reached. Stathis [6] outlined the defect generation mechanisms at low voltage in

Fig. 2.3. At least three defect generation mechanisms have been identified: the thermochemical model (E-model), the anode hole injection (AHI) model, and the anode hydrogen release (AHR) model. All three models are briefly discussed below.

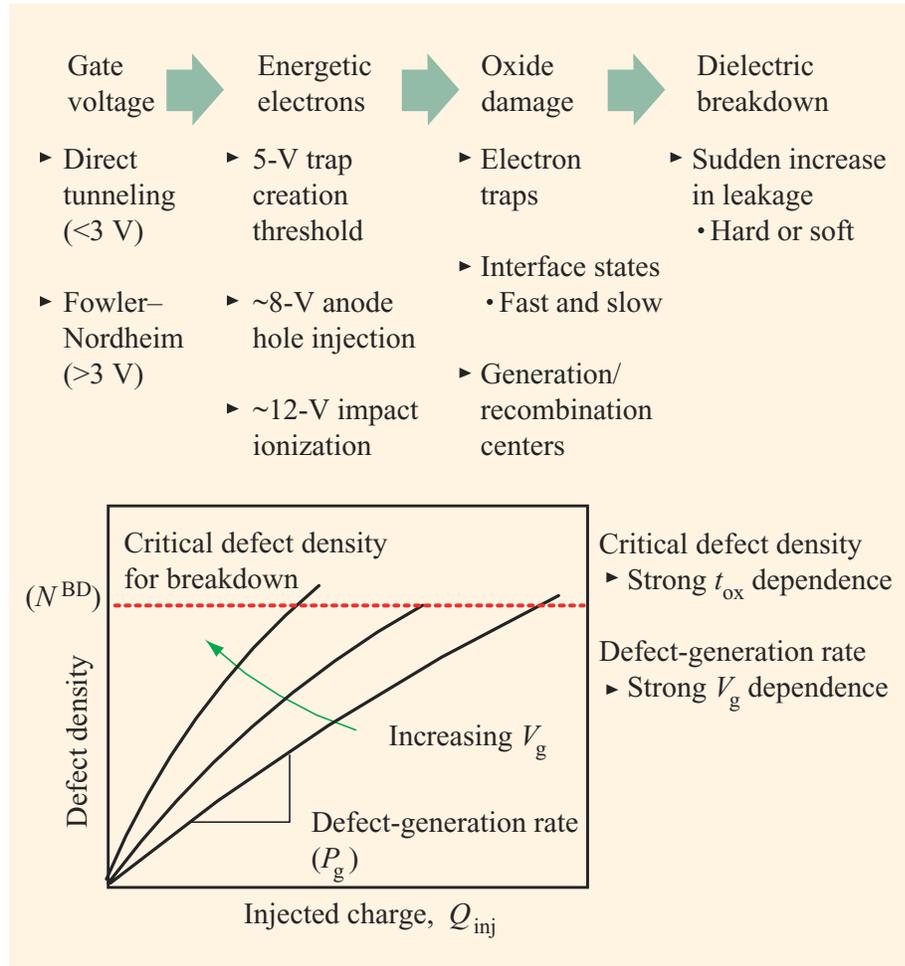


Figure 2.3: Outline of the mechanism of defect generation leading to breakdown in SiO_2 [6].

2.2.3.1 Thermochemical Model

The thermochemical model (**E**-model) indicates that the applied electric field interacts with the weak $Si-Si$ bonds associated with oxygen vacancies in the amor-

phous SiO_2 film. Tunneling electrons are not necessary (to the first order) in the thermochemical model to create defects [32]. McPherson [36] reviewed the development of this model and proposed a physical explanation. This model proposes that defect generation is a field-driven process and the current flowing through the oxide plays at most a secondary role. The interaction of the applied electric field with the dipole moments associated with oxygen vacancies (weak $Si-Si$ bonds) in SiO_2 lowers the activation energy required for thermal bond breakage and accelerates the dielectric degradation process. Eventual charge trapping at those broken bond sites and their wave function overlap lead to a conduction subband formation and severe Joule heating at the stage of oxide breakdown. McPherson [36] also showed that allowing for a distribution of energies of the weak bonds could account for a wide range of observations of the temperature and field dependence of dielectric breakdown times. The **E**-model suggests T_{BD} is given by

$$T_{BD} = A_0 \exp(-\gamma \mathcal{E}_{ox}) \exp\left(\frac{E_a}{kT}\right), \quad (2.2)$$

where:

A_0 : arbitrary scale factor, dependent upon materials and process details,

γ : field acceleration parameter,

\mathcal{E}_{ox} : externally applied electric field across the dielectric.

The **E**-model has attained widespread acceptance on the basis of experimental verified exponential dependence of T_{BD} on field [37]. But this is not enough to prove the validity of this model. It was observed that for very thin oxides the breakdown

times are no longer a function of only the field, but strongly decreased with thickness at the same oxide field. The decreasing breakdown times are consistent with the increasing direct-tunneling leakage currents in the ultrathin oxides. An AHI-like mechanism was proposed that the strong increase in current leads to increase in hole injection, and that these holes are trapped at oxygen vacancies further reducing the activation energy for bond rupture. Substrate-hot-electron (SHE) injection experiments showed that this T_{BD} is inversely related to the current density, showing that breakdown is dominated by the effect of the energetic electrons and not the field in the oxide [38].

2.2.3.2 AHI Model

The AHI model ($\mathbf{1}/\mathbf{E}$ -model) was proposed by Schuegraf and Hu [39]. It claims that breakdown is caused by holes which are injected from the anode contact. Electrons injected from the gate metal cathode into the oxide undergo impact ionization events that generate holes in the process. These holes become trapped in the oxide near the cathode, distorting the band diagram and increasing the field nearby as shown in Fig. 2.4. Electron tunneling is enhanced in the high field according to Fowler-Nordheim tunneling Eq. (2.1), thus resulting in greater current injection. Another mechanism is that at the anode side of the oxide the electron drops down to Fermi level and may donate its energy of at least 3.1 eV to the lattice at the SiO_2 -metal interface [40]. This energy is sufficient to break an $Si - O$ bond. The breaking of bonds proceeds from anode to cathode and forms a convenient conductive path

for discharge that causes a dielectric breakdown. In both cases the injected oxide charge is accumulated inside the oxide until a critical hole charge density is reached for the dielectric breakdown.

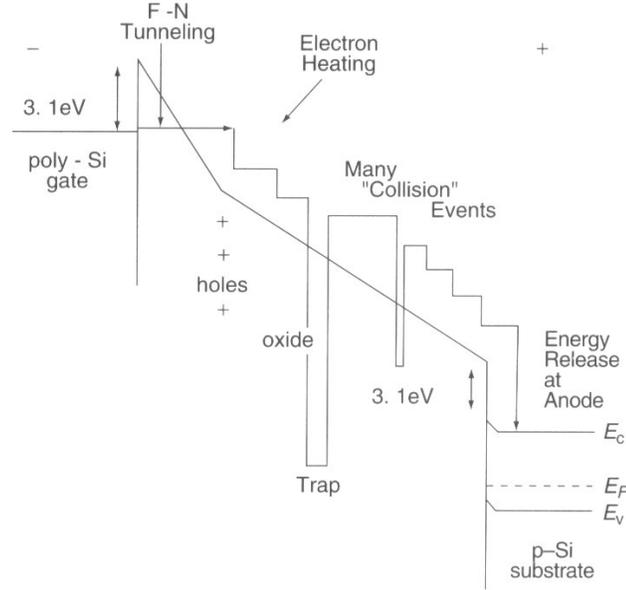


Figure 2.4: Band diagram models of AHI model [4].

The reciprocal field expression of time-to-breakdown (T_{BD}) based on AHI model takes the form

$$T_{BD} = \tau_0(T) \exp\left(\frac{G(T)}{\mathcal{E}_{ox}}\right), \quad (2.3)$$

where \mathcal{E}_{ox} is the electric field across the dielectric in MV/cm. Constants $\tau_0(T)$ and $G(T)$ are temperature-dependent and given by $\tau_0(T) = 5.4 \times 10^{-7} \exp[-0.28eV/kT](\text{sec})$, and $G(T) = 120 + \frac{5.8}{kT}$ MV/cm, where k is Boltzmann's constant and T is the absolute temperature [41].

Support of the AHI model comes from a theoretical treatment of anode hole injection by surface plasmon excitations, and experimental data showing the expected

dependence on anode material [37]. This model was criticized for its inability to account for the substrate currents measured at lower voltages since the gate voltage threshold for positive charge generation by hole trapping due to AHI is 7-8 V according to the plasmon model [42]. The AHI model has been criticized because there are other origins of substrate current at low voltage beside tunneling holes, which include the generation-recombination process in the substrate [43], photo-excitation due to photons generated by hot electrons in the gate [44]. However, some recent experimental evidence and modeling have demonstrated the effectiveness of the AHI model in ultrathin oxides at low gate voltage [32].

2.2.3.3 AHR Model

There is evidence for an AHR model that involves the release of atomic hydrogen from the anode by energetic tunneling electrons [45]. The released hydrogen diffuses through the oxide and can generate electron traps. Experiments have shown that exposure of bare SiO_2 films to atomic hydrogen radicals, even without any electric field, will produce electrically active defects essentially identical to those produced by electrical stress or radiation [37]. DiMaria [45] showed the desorption rate of hydrogen from Si surfaces is similar to the voltage dependence of the trap generation process. Based on data taken at IBM, he determined that hydrogen release requires electrons with energy levels of at least 5 eV in the anode, and 2 eV in the oxide. The trap-creation process continues in the subthreshold region even at operating voltages down to 1.2 V or lower [46].

The primary argument against the hydrogen release process for oxide breakdown is the apparent lack of any isotope effect for the breakdown process when compared to the large effect observed for hydrogen/deuterium desorption and for channel hot electron induced interface degradation [47]. The observation of T_{BD} does not appear to improve if an isotope of hydrogen is used to passivate the silicon-oxide interface [48].

Degraeve et al. [5] gave an outline of these three models on neutral electron trap generation as shown in Fig. 2.5. Fig. 2.5 (a) shows the thought of the AHI model. The thermochemical model is shown in Fig. 2.5 (c) and the AHR model is included in “Other Mechanism” of Fig. 2.5 (b).

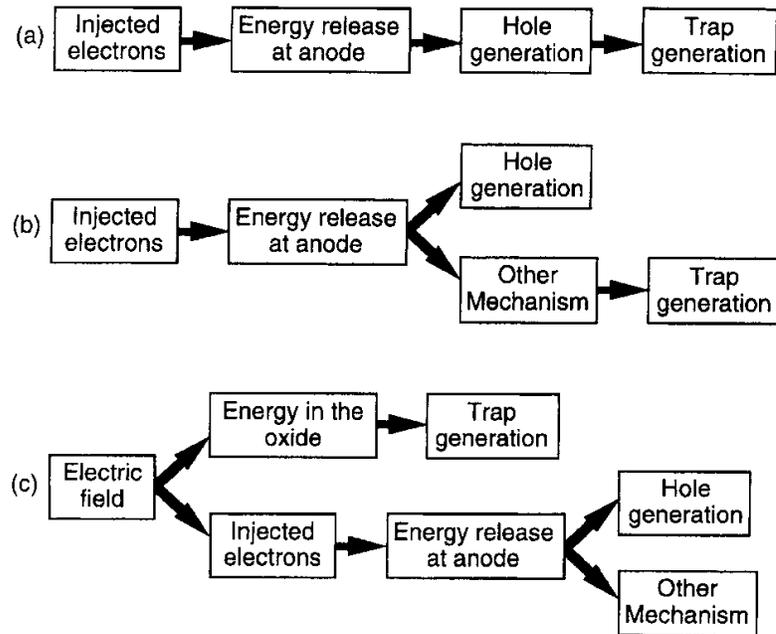


Figure 2.5: Outline of neutral electron trap generation [5].

2.2.3.4 Voltage Dependence of Defect Generation

The **E**-model and the **1/E**-model both proposed that the electric oxide field is the root cause of dielectric breakdown. They were used to model thick/thin oxide breakdown and both described some experimental data quite well [49, 50]. Later, the **E**-model gained popularity and the **1/E**-model was generally accepted to be only appropriate at high voltages in thick oxides. However, as the oxide thickness reached into the ultrathin range where direct tunneling of electrons dominates, the **E** model was experimentally shown not appropriate to model ultrathin oxide breakdown [51]. Below the 5V threshold for FN tunneling, the defect generation rate depends only on the absolute value of the gate voltage [16], the oxide breakdown is voltage driven [52, 51, 53].

Wu et al. [54] proposed a power-law voltage-dependence model ($T_{BD} \sim V_g^{-n}$) from a variety of ultrathin oxides experiments. The power law model was further experimentally verified by [55, 56, 57].

2.3 TDDB Models

2.3.1 Lifetime Models

2.3.1.1 Voltage/Field Dependence

For thick oxides (> 4 nm) stressed at high voltages in non-ballistic FN regime, both the **E**-model and **1/E**-model can be applied. However, the **E**-model has gained widespread acceptance because the logarithm of T_{BD} was reported to be linear with

electric fields closer to operating conditions [58, 59, 60, 61], including an experiment of three years' duration at the oxide field down to 5.3MV/cm on 9-nm films [62].

The TDDB lifetime dependence on field are:

- **E**-model:

$$T_{BD} \propto \exp(-\gamma \cdot \mathcal{E}_{OX}). \quad (2.4)$$

- **1/E**-model:

$$T_{BD} \propto \exp\left(\frac{\gamma}{\mathcal{E}_{OX}}\right). \quad (2.5)$$

The electric oxide field can be substituted by the gate voltage as a work hypothesis.

For ultrathin oxides (< 4 nm) stressed in the ballistic FN tunneling or direct tunneling regimes, the power law model is proposed to model the voltage driven TDDB.

$$T_{BD} \propto V_g^{-n}. \quad (2.6)$$

The reported exponent n values are listed in Table 2.1.

2.3.1.2 Temperature Dependence

Generally it was assumed that TDDB temperature dependence follows the Arrhenius relationship:

$$T_{BD} \propto \exp\left(\frac{E_a}{kT}\right), \quad (2.7)$$

where E_a is the activation energy. In many early studies, E_a was observed to decrease for increasing applied gate voltage or electric field [60, 65, 66], and was also observed to change with temperature [67]. For ultrathin oxides, a steeper

n	t_{ox} (nm)	V_g (V)	T (°C)	Device	Ref
44	1.2-5.0	2-6	140	capacitor, NFET and PFET	[63]
41.7	1.6-2.2	2.7-4.0	140	NFET	[57]
44	3.0-12	3.5-5.3	140	NFET	[64]
30	3.0-12	5.3-8.0	140	NFET	
45	1.6-3.0		125	NFET (inversion)	[55]
40				NFET (accumulation)	
44				PFET (accumulation)	
45		$ V_g > 3.8$		PFET (inversion)	
38		$ V_g < 3.8$		PFET (inversion)	

Table 2.1: Reported values of the power law exponent n .

temperature dependence was reported as compared with thick oxides [68, 69, 70, 71]. The stronger dependence may be due to the lower voltages used to stress thinner oxides. More important, non-Arrhenius temperature dependence has been reported. Considering the activation energy's dependence on temperature and voltage, Wu et al. [72] suggested a new temperature dependence model.

$$T_{BD} \propto \exp\left(\frac{a(V)}{T} + \frac{b(V)}{T^2}\right), \quad (2.8)$$

where the prefactor a and b are voltage-dependent. b/T^2 is included for possible non-Arrhenius temperature effect.

2.3.1.3 Area Dependence

Gate oxide breakdown is a weakest-link type of failure because the whole chip fails if any one device fails, and a device fails if any small portion of the gate area of the device breaks down. Stathis [17] derived the area dependence of TDDB lifetime:

$$T_{BD} \propto \left(\frac{1}{A}\right)^{1/\beta}, \quad (2.9)$$

where A is the area of oxide, β is the shape parameter of Weibull distribution. The area scaling is helpful in relating breakdown tests on individual small area capacitors to the reliability of an integrated circuit containing many millions of gates.

2.3.2 Statistical Models

2.3.2.1 Percolation Theory

The percolation theory was applied to modeling the intrinsic breakdown distribution by Degraeve et al. [73]. Stathis [17] further applied this theory in simulation to demonstrate the thickness dependency of the number of defects at breakdown.

Fig. 2.6 shows the percolation model for oxide breakdown. It is assumed that electron traps are generated inside the oxide at a random position in space. Around these traps a sphere is defined with a fixed radius r , which is the only parameter of this model(Fig. 2.6(a)). If the spheres of two neighboring traps overlap, conduction between these traps becomes possible. The two interfaces are modeled as an infinite set of traps (Fig. 2.6(b)). This mechanism of trap generation continues until a conducting path is created from one interface to the other and breakdown happens.

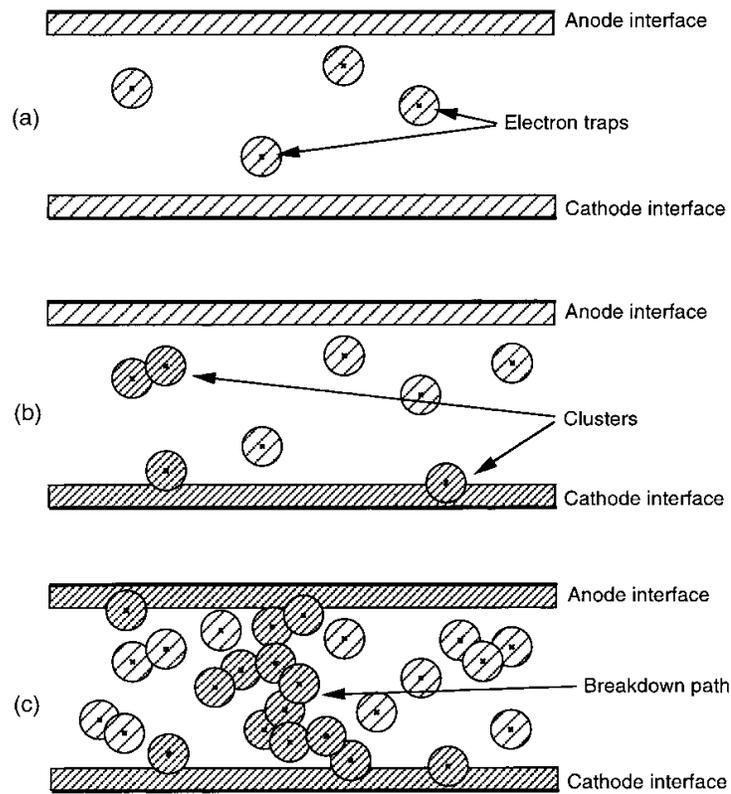


Figure 2.6: The percolation model for oxide breakdown [5].

The percolation model for breakdown is able to explain quantitatively two important experimental observations: (1) as the oxide thickness decreases, the density of oxide traps needed to trigger breakdown decreases; and (2) as the oxide thickness decreases, the Weibull slope of breakdown distribution decreases [5].

2.3.2.2 Weibull Distribution

The statistics of gate oxide breakdown are usually described using the Weibull distribution

$$F(t) = 1 - \exp[-(t/\alpha)^\beta], \quad (2.10)$$

where α is the scale parameter (63.2 percentile) and β is the shape parameter. Weibull distribution is an extreme-value distribution in $\ln(x)$ and is a ‘weakest link’ type of problem. Here F is the cumulative failure probability, x can be either time or charge, α is the scale parameter and β is the shape parameter. The ‘weakest link’ model was formulated by *Suñé et al.* [74] and described oxide breakdown and defect generation via a Poisson process. In this model, a capacitor is divided into a large number of small cells. It is assumed that during oxide stressing neutral electron traps are generated at random positions on the capacitor area. The number of traps in each cell is counted, and at the moment that the number of traps in one cell reaches a critical value, breakdown will occur. *Dumin* [58] incorporated this model to describe failure distributions in thin oxides.

The Weibull slope β is an important parameter for reliability projections. A key advance was the realization that β is a function of oxide thickness t_{ox} , becoming

smaller as t_{ox} decreases [17, 75, 18]. The smaller β for thinner oxide is explained as the conductive path in the thinnest oxides consists of only a few traps and therefore has a larger statistical spread. The shape parameter's oxide thickness dependence is shown in Fig. 2.7. The dependence of β on t_{ox} can be fitted to

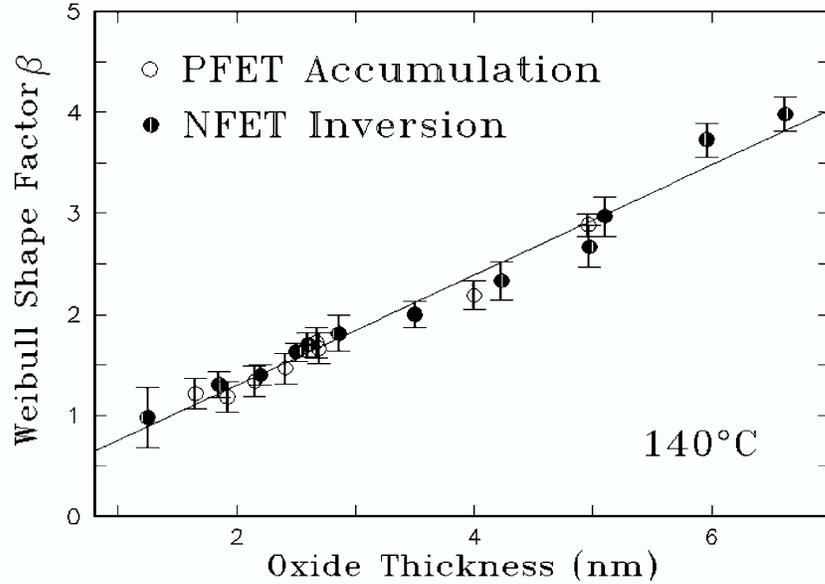


Figure 2.7: Weibull shape parameters vs. oxide thickness [7].

$$\beta = (t_{int} + t_{ox})/a_0, \quad (2.11)$$

where a_0 is the linear defect size with a fitted value of 1.83 nm, and t_{int} is the interfacial layer thickness with a fitted value of 0.37 nm [76]. It can be found that β is approaching one as T_{ox} is near 1 nm, which means Weibull distribution will become exponential distribution.

Log-normal distribution has also been used to analyze accelerated test data of dielectric breakdown. Although it may fit failure data over a limited sample set, it has been demonstrated that the Weibull distribution more accurately fits

large samples of TDDB failures [77]. An important disadvantage of log-normal distribution is that it does not predict the observed area dependence of T_{BD} for ultrathin gate oxides.

Chapter 3

Negative Bias Temperature Instability

3.1 Introduction

Negative bias temperature instability (NBTI) happens to PMOS devices under negative gate voltages at elevated temperatures. The degradation of device performance, mainly manifested as the absolute threshold voltage V_{th} increase and mobility, transconductance and drain current $I_{d_{sat}}$ decrease, is a big reliability concern for today's ultra-thin gate oxide devices. NBTI has been studied and modeled since the 1960s [19]. Deal [78] named it "Drift VI" and discussed the origin in the study of oxide surface charges. Goetzberger et al. [79] investigated surface states change under combined bias and temperature stress through experiments that utilized MOS structures formed by a variety of oxidizing, annealing and metalizing procedures. They found an interface trap density D_{it} peak in the lower half of the band gap and p-type substrates gave higher D_{it} than n-type substrates. The higher the initial D_{it} , the higher the final stress-induced D_{it} . Jeppson et al. [8] first proposed a physical model to explain the surface trap growth of MOS devices subjected to negative bias stress. The surface trap growth was described as diffusion controlled at low fields and tunneling limited at high fields. The power law relationship ($t^{1/4}$) was also proposed for the first time. The study of NBTI has been very active in recent years since the interface trap density induced by NBTI increases with decreasing

oxide thickness which means NBTI is more severe to ultra-thin oxide devices. New developments of NBTI modeling and surface trap analysis have been reported in recent years. At the same time, effects of various process parameters on NBTI had been studied in order to minimize the NBTI. Schroder et al. [21] reviewed pre-2003 experimental results and various proposed physical models together with the effects of manufacturing process parameters. Detailed latest reviews can also be found in [80, 81, 82, 83, 84]. In this section, the up-to-date research discoveries of NBTI failure mechanism, models and related parameters will be briefly discussed.

3.2 Physics of Failure

Silicon dioxide, the critical component of silicon devices, serves as insulation and passivation layers and are never completely electrically neutral. Mobile ionic charges, oxide trapped electron or holes, fabrication-process-induced fixed charges and interface trapped charges are four main categories of charges inside oxide and at the silicon-oxide interface. The electrical characteristics of a silicon device are very sensitive to the density and properties of those charges. As already known, the threshold voltage of PMOSFET is given by

$$V_{th} = V_{FB} - 2\phi_B - |Q_B|/C_{ox}, \quad (3.1)$$

where $\phi_B = (kT/q)\ln(N_D/n_i)$, $|Q_B| = (4q\epsilon_{Si}\phi_B N_D)^{1/2}$ and C_{ox} is the oxide capacitance per unit area. The flat band voltage V_{FB} is given by

$$V_{FB} = \phi_{MS} - \frac{Q_f}{C_{ox}} - \frac{Q_{it}(\phi_s)}{C_{ox}}, \quad (3.2)$$

where Q_f is the fixed charge and Q_{it} the interface trapped charge. From Eq. (3.1) and Eq.(3.2), it can be found that the only parameters to change the threshold voltage are the Q_f and Q_{it} . Most research works of NBTI failure mechanism have been focused on the generations of Q_{it} and Q_f .

3.2.1 Interface Trap Generation: Reaction-Diffusion Model

The Reaction-Diffusion (R-D) model is the most prevalent among various proposed NBTI models. Jeppson and Svensson [8] were the first to propose the R-D model to explain the generation of interface states at low fields. In this model, it is assumed that the silicon interface contains a large number of defects that are electrically inactive and can be activated through chemical reaction like this



$X_{interface}$ is a diffusing species which is formed at the interface in the reaction. Based on the infrared measurements report which showed large numbers of $S_i - H$ groups existing in bulk silicon and probably also at the interface, Jeppson et al. proposed this reaction



where $Si_3 \equiv SiH$ is the surface defect, $Si_3 \equiv Si \cdot$ is the surface trap, $O_3 \equiv Si^+$ is the oxide charge and $O_3 \equiv SiOH$ is the diffusing X . When the defect is activated, the H of SiH bond is released by some dissociation mechanisms and reacts with the

SiO_2 lattice to form an OH group bonded to an oxide atom, leaving a trivalent Si atom in the oxide to form a fixed charge and one trivalent Si atom at the Si surface to form an interface trap. This chemical reaction is schematically shown in Fig 3.1. The $N_{it} \sim t^{1/4}$ relationship was observed and mathematically proved by assuming the process is diffusion limited rather than reaction-rate limited.

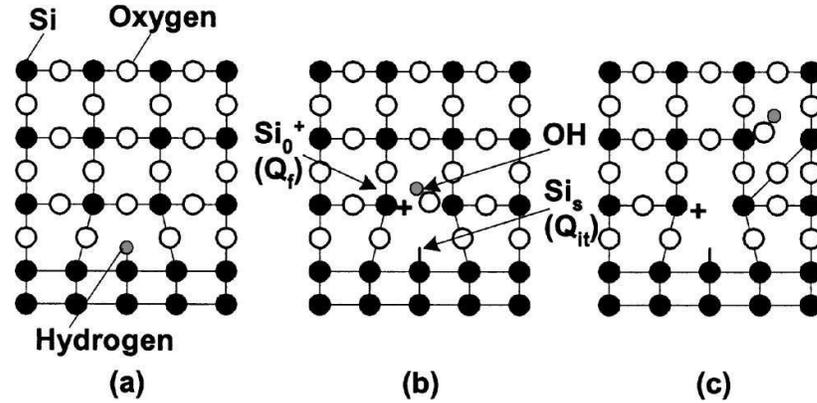


Figure 3.1: Schematic two-dimensional representation of the $Si - SiO_2$ interface, showing (a) the $\equiv Si$ defect, (b) how this defect may be electrically activated during NBTI to form an interface trap, a fixed oxide charge, and a hydroxyl group, and (c) the OH diffuses through the oxide. Adapted from [8].

Various mechanisms have been proposed for the dissociation process. Ogawa et al. [85] listed three of those:

1. High-electric field dissociation

$Si_3 \equiv SiH \longrightarrow Si_3 \equiv Si \cdot + H_i$, where H_i , the neutral species X, is an interstitial hydrogen atom.

2. Interstitial atomic hydrogen attack

$Si_3 \equiv SiH + H_i \longrightarrow Si_3 \equiv Si \cdot + H_2$, molecular hydrogen H_2 is the species X.

3. Dissociation involves holes



The actual diffusing species X have not yet been identified. Possibilities include interstitial atomic hydrogen (H_i) [85], molecular hydrogen (H_2), hydroxyl (OH) group and proton (H^+) [86]. Rashkeev et al. [86] did first-principles calculations to show that the proton is the only stable charge state of H at the $Si - SiO_2$ interface. The protons can react directly with SiH to form H_2 and leave behind positively charged dangling bonds. Alam et al. [84] proposed that the H is released as atomic H , then convert to and diffuse as molecular H_2 , and the measurement delay is the main reason of various diffusion species observations.

3.2.2 Fixed Charge Generation

The fixed charge Q_f is a positive charge in the oxide and near the $Si - SiO_2$ interface. It can't be charged or discharged by varying the silicon surface potential. It is primarily due to excess silicon species introduced during oxidation and during post-oxidation heat treatment [19]. Negative bias stress can also increase its value like the generation of interface trap. Ogawa et al. [85] determined fixed oxide charge densities from capacitance-voltage measurements and interface trap densities from conductance measurements of MOS capacitors under low fields stress (-1.6 to -5.0 MV/cm). The formulated expressions for N_f and N_{it}

$$\Delta N_f(E_{ox}, T, t) = A_1 \mathcal{E}_{ox}^{1.5} t^{0.14} \exp(-0.15/kT), \quad (3.6)$$

$$\Delta N_{it}(E_{ox}, T, t, t_{ox}) = A_2 \mathcal{E}_{ox}^{1.5} t^{0.25} \exp(-0.2/kT)/t_{ox}, \quad (3.7)$$

where A_1 and A_2 are two constants that are independent of \mathcal{E}_{ox} , T , and t_{ox} . The thickness of oxide in their experiments is ranged between 4.2-30 nm, together with an early report [19] which stated no thickness dependence of fixed charges for 40-100 nm oxides, and showed that ΔN_f is independent of oxide thickness in a wide range. ΔN_{it} is inversely proportional to oxide thickness as shown in Eq. 3.7, which means NBTI is worse for thinner oxides.

3.2.3 Recovery and Saturation

Another important phenomena of NBTI is the recovery of the threshold voltage shift after the negative bias stress is removed [8]. This means NBTI may have different characteristics between DC and AC operation. Abadeer et al. [87] reported a 3X increase in the magnitude of threshold voltage shift under DC operation than that of AC operation. Rangan's experiment [88] showed the recovery is independent of stress voltage, time and temperature (under 25 °C) and can reach 100% at 25 °C for gate oxides ranging from 4.5 to 15.0 nm. The mechanisms of recovery is still under investigation. One explanation is that the diffusion species X moves back to the $Si - SiO_2$ interface under the influence of positive gate voltage and passivates the Si dangling bond [89]. Another interpretation is the delicate interplay between forward dissociation and reverse annealing rates during the stress and relaxation phases of AC degradation [90]. At each stress interval the V_{th} degradation at first returns quickly, then continues to degrade more slowly. The ratio of AC to DC degradation is affected by the duty cycle. It was reported that the degradation

under AC operation has little or no frequency dependence up to 500KHz [19, 89, 91] but then decreases further above 2MHz [87].

There were reports that indicated that NBTI shifts tend to saturate over time [21, 92, 93]. One possible reason is the reaction limitation mechanism. The generation of Si^+ decreases as the number of available SiH bonds reduces with time. Another possible reason is if the diffusing species encounters a new interface at which it is not transferred across but reflected [93]. Saturation may have important implications for long term reliability prediction [94, 95]. However, no agreement of the physical understanding behind the saturation has been reached. Alam et al. [84] proposed that saturation is an artifact of measurement delay.

3.3 NBTI Models

3.3.1 Degradation Models

The time dependence of the threshold voltage shift (ΔV_{th}) is found to follow a power-law model

$$\Delta V_{th}(t) = At^n, \quad (3.8)$$

where A is a constant which depends on oxide thickness, field and temperature. The time-exponent n is a sensitive measure of the diffusion species. The theoretical value of the exponent parameter n is 0.25 according to the solution of diffusion equations [8]. Chakravarthi [93] suggested that n varies around 0.165, 0.25 and 0.5 depending on the reaction process and the type of diffusion species. According to Alam et al. [84], $n = 1/2$ for proton, $n = 1/6$ for molecular H_2 , $n = 1/4$ for atomic H . n was

also reported to change from ~ 0.25 initially (stress time ~ 100 s) to 0.16 at 10^6 s stress time [82].

NBTI degradation is thermally activated and sensitive to temperature. The temperature dependence of NBTI is modeled by the Arrhenius relationship. The activation energy appears to be highly sensitive to the types of potential reacting species and to the type of oxidation methods used [21]. Reported activation energies range from 0.18-0.84 eV [96, 97].

Improved models have been proposed after the simple power-law model. Considering the temperature and gate voltage, ΔV_{th} can be expressed as

$$\Delta V_{th}(t) = B \exp(\beta V_g) \exp(-E_a/kT) t^{0.25} \quad (3.9)$$

where B and β are constants and V_G is the applied gate voltage. Considering the effects of gate voltage and oxide field, Mahapatra et al. [98] proposed a first order N_{it} model.

$$\Delta N_{it}(t) = K(C_{ox}(V_g - V_{th}))^{0.5} \exp(\beta \mathcal{E}_{ox}) \exp(-E_a/kT) t^{0.25}, \quad (3.10)$$

where K is a constant.

3.3.2 Lifetime Models

NBTI failure is defined as ΔV_{th} reaches a threshold value. Based on the degradation models such as Eq 3.7 and 3.9, NBTI lifetime can be represented as:

1. Field model.

$$\tau = C_1 \mathcal{E}_{ox}^{-n} \exp(E_a/kT), \quad (3.11)$$

where C_1 is a constant.

2. Voltage model.

$$\tau = C_2 \exp(-\beta V_g) \exp(E_a/kT), \quad (3.12)$$

where C_2 is a constant.

3.3.3 Statistical Model

Little work has been done to model the NBTI lifetime statistically with significant sample size. Lognormal distribution has been used for general purpose.

Chapter 4

Hot Carrier Degradation

4.1 Introduction

Hot carrier degradation (HCD) has been studied for more than 30 years as an important failure mechanism that must be handled in design of aggressively scaled VLSI devices. Extensive work has been done to discover the physical mechanisms, model lifetime distribution and improve technology. Physical understanding of HCD and models are briefly discussed.

4.2 Physics of Failure

Semiconductors in thermal equilibrium has an average energy gain of zero as electrons and holes continually absorb and emit acoustical phonons (low-frequency lattice vibrations). If the electrical field is very high, for an example, 1MV/cm, the carriers gain more energy than they lose by scattering. Such accelerated electrons have energies of $E_c + kT_e$, where E_c is the conduction band edge, T_e is an effective temperature. With effective temperatures ($\sim E_c/kT$) of tens of thousands of degrees Kelvin, these electrons are at the very top of the Fermi distribution—known as *hot electron* [4].

In the operation of MOSFET, if the gate voltage is comparable to or lower

than V_{ds} , the inversion layer is much stronger on the source side than the drain side (if $V_d > V_s$) [9], and the voltage drop due to channel current is concentrated on the drain side. The field near the drain side can be so high that carriers can gain enough energy between two scattering events to become hot carriers. The majority of these hot carriers simply continues toward the drain, but a small number of them gain enough energy to generate electrons and holes by impact ionization. For NMOSFET, the vast majority of the generated holes are collected by the substrate and give rise to the substrate current (I_{sub}), and the generated electrons enhance the drain current (I_d). Photo emission may also happen during hot carrier generation in the drain.

Some of the hot carriers with enough energy (about 3.1 eV for electrons and 4.6 eV for holes) [9] can surmount the energy barrier at the $Si - SiO_2$ interface and be injected into the oxide, producing a small gate current (I_g). Some energetic injected carriers may break some $Si - H$ or similar weak bonds in the oxide or at the SiO_2 interface. If the hot carrier injection lasts long enough, the trapped charge or generated defects will permanently modify the electric field at the $Si - SiO_2$ interface and hence the electrical characteristics of the MOSFET such as channel mobility, threshold voltage and drain current. Fig. 4.1 schematically shows the process of HCD. In general, hot electron injection is much more likely to happen because electrons have smaller effective mass and the $Si - SiO_2$ interface energy barrier is larger for holes (≈ 4.6 eV) than for electrons (≈ 3.1 eV). NMOS suffers HCD more than PMOS.

According to Takeda [99], there are three main types of hot carrier injection

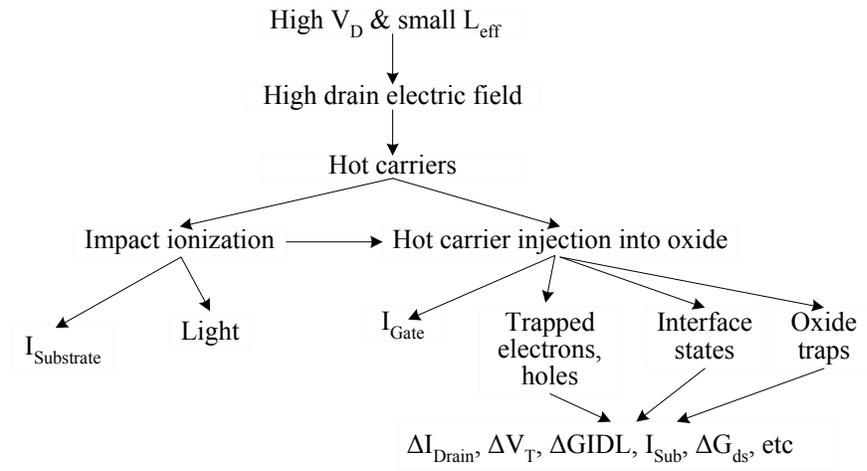


Figure 4.1: Hot carrier generation and degradation in MOSFETs. Adapted from [9].

modes:

1. Channel hot electron (CHE) injection,
2. Drain avalanche hot carrier (DAHC) injection,
3. Secondary generated hot carrier (SGHE) injection.

Fowler-Nordheim tunneling and direct tunneling may also cause hot carrier injection [100], especially at deep sub-micrometer level.

4.2.1 Channel Hot Electron (CHE) Injection

The CHE injection is the dominating injection mechanism at stress conditions $V_d/2 < V_g \leq V_d$, especially for short-channel NMOSFETs. Depicted in Fig. 4.2 is an NMOSFET where CHE injection occurs when the gate voltage (V_g) is comparable to

the drain voltage (V_d). The gate current (I_g) rises as V_g initially increases and peaks when V_g is roughly equal to the drain-source potential V_d , and drops thereafter. There are two reasons that cause the I_g to increase. First, the inversion charge in the channel increases, so that more electrons are present for injection into the oxide. Second, the stronger influence of the vertical electric field in the oxide attracts more electrons and prevents them from de-trapping and drifting back into the channel. Beyond $V_g \approx V_d$, further increase of V_g leads to a more severe reduction in the lateral field and decrease of the gate current [101]. It was reported [102] that, if an n-channel MOSFET is operating at $V_g = V_d$, the conditions would be optimum for CHE injection of “lucky electrons.” Such electrons gain sufficient energy to surmount the $Si - SiO_2$ barrier without suffering an energy-losing collision in the channel. In many cases, this gate current is responsible for device degradation as a result of carrier trapping. No gate current can be measured for $V_g < V_d$, since CHE injection is retarded. However, if V_d is large, reduction of V_g intensifies the electric field at the drain to the point where avalanche multiplication due to impact ionization may substantially increase the supply of both hot electrons and hot holes.

4.2.2 Drain Avalanche Hot Carrier (DAHC) Injection

The DAHC injection occurs around a maximum substrate current condition or at $V_g = V_d/2$ in NMOSFETs and in deep-submicron PMOSFETs [103]. It is schematically shown in Fig. 4.3. This mechanism first depends on an impact-

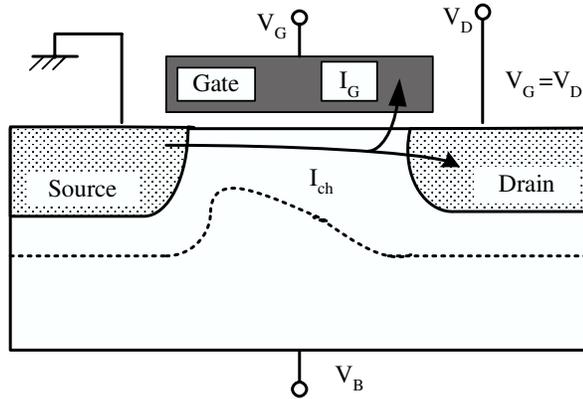


Figure 4.2: Channel-hot-electron injection. Occurs when $V_g \approx V_d$.

ionization avalanche to create carriers. These secondary carriers inject into the oxide around the drain edge and recombination takes place. Then interface traps are generated by some mechanism breaking $Si-H$ bonds at the $Si-SiO_2$ interface such as direct energy transfers and hydrogen releases [104]. Analyzing DAHC behavior is difficult because hot holes and hot electrons are injected simultaneously into the oxide and across the drain junction just below the substrate surface.

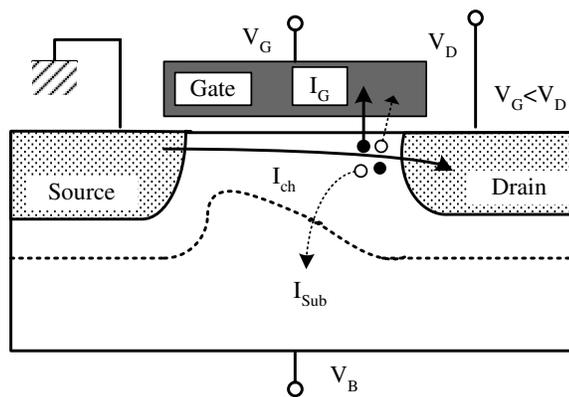


Figure 4.3: Drain avalanche hot-carrier injection. Occurs when $V_d > V_g$.

4.2.3 Secondary Generated Hot Electron (SGHE) Injection

SGHE involves the generation of hot carriers from impact ionization involving a secondary carrier that is likewise created by an earlier incident of impact ionization. Secondary impact ionization by hot holes and photoinduced generation processes have been reported as secondary minority carrier generation mechanisms (as shown in Fig. 4.4). Takeda [99] experimentally showed photoinduced generation is the main mechanism. The temperature dependence of I_{sub} and that of electron diffusion current, I_d , was compared for a device with $t_{ox} = 7$ nm and $L_{eff} = 2.0\mu m$. The experiment results imply that a photoinduced generation process, believed to be bremsstrahlung radiation, rather than secondary impact ionization, is more likely to be the origin of the SGHE.

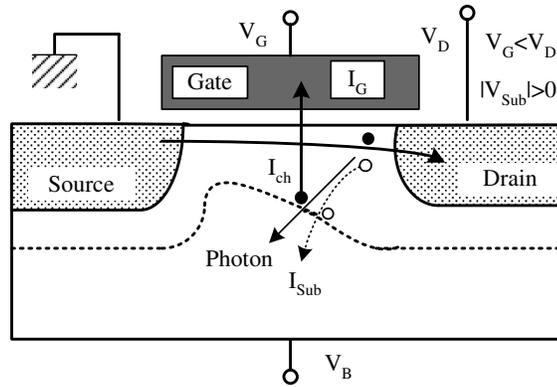


Figure 4.4: Secondarily generated hot electron injection.

4.3 HCD Models

4.3.1 Lucky Electron Model (LEM)

The widely used lucky electron approach of modeling the hot electron distribution was originated by Shockley [105]. Verwey [106] applied it in the study of substrate hot electron injection that was subsequently refined and verified by Ning [107]. Hu [108, 109] modified the substrate lucky electron injection model and applied to CHE in NMOSFET under DC stress conditions.

The basic assumption of the lucky electron model hinges on there being a supply of hot electrons that are “lucky.” In order for channel electrons to reach the gate oxide, two lucky processes are involved. The first requires that electrons gain sufficient kinetic energy from the channel field to become “hot.” Secondly, the electron momentum must be redirected perpendicularly, so that hot electrons can enter the oxide. The probability that a channel electron will travel a distance d or more without suffering any collision is equal to $\exp[-d/\lambda]$, where λ is the mean free path between scattering events. Consider an electron of charge q , traveling a distance λ in the channel electric field \mathcal{E}_c . The probability that it will reach energy ϕ without suffering a collision is given by $\exp[-\phi/(\lambda q \mathcal{E}_c)]$, as $d = \phi/(q \mathcal{E}_c)$. The probability of a hot electron’s redirection to the $Si - SiO_2$ interface without suffering any collision is essentially a function of oxide field \mathcal{E}_{ox} [107]. The explicit consequences of these processes are substrate current (I_{sub}) and gate current (I_g) whose magnitudes depend on electron energies reaching those required for impact ionization (ϕ_i), and for surmounting the $Si - SiO_2$ energy barrier (ϕ_b), respectively.

Hu [109] presented a general model for the hot electron effects. The I_{sub} , I_g , device lifetime (τ) and the change of interface traps (ΔN_{it}) are modelling as followed.

$$I_{sub} = C_1 I_d \exp\left(-\frac{\phi_i}{q\lambda\mathcal{E}_m}\right), \quad (4.1)$$

$$I_g = C_2 I_d \exp\left(-\frac{\phi_b}{q\lambda\mathcal{E}_m}\right), \quad (4.2)$$

$$\Delta N_{it} = C_3 \left[t \frac{I_d}{W} \exp\left(-\frac{\phi_{it}}{q\lambda\mathcal{E}_m}\right) \right]^n, \quad (4.3)$$

$$\tau = C_4 \frac{W}{I_d} \exp\left(\frac{\phi_{it}}{q\lambda\mathcal{E}_m}\right), \quad (4.4)$$

All the parameters are listed below.

- I_d is the drain current flow that supplies some of the eventually lucky electrons.
- W is the channel width.
- ϕ_i is the minimum energy for electron to create an impact ionization (reported ϕ_i varied from 1.2 eV to almost 3 eV [109]).
- ϕ_b is the $Si - SiO_2$ interface barrier energy ($\phi_b = 3.1$ eV).
- ϕ_{it} is interface trap creation energy for electron with an estimated value of 3.7 eV [109].
- λ is the hot electron mean-free-path, which is temperature-dependent. Several different values of λ have been reported (9.2 nm [107], 8.9 nm [110] 7.8 nm [111] and 6.7 nm [112]).

- n in Eq. 4.3 ranges from 0.5 to 1 as the interface trap generation process is similar to the rate of thermal oxidation. t is time.
- C_1 is constant, which is a weak function of \mathcal{E}_m , and the device parameters [109], C_2 and C_3 are process-related constants. C_1 , C_2 and C_3 need to be characterized for each technology node. C_4 is a constant related to failure criteria.
- \mathcal{E}_m , the maximum channel electric field, is the most important parameter in Eq. (4.1), (4.2), (4.4) and (4.3).

A semi-quantitative approximation \mathcal{E}_m model has been given in [109]:

$$\mathcal{E}_m = \frac{V_d - V_{dsat}}{\sqrt{3t_{ox}x_j}}, \quad (4.5)$$

where t_{ox} is the gate oxide thickness, x_j is the drain junction depth. $\sqrt{3t_{ox}x_j}$ is the approximate effective “pinchoff” region length. The factor 3 in $\sqrt{3t_{ox}x_j}$ derives from the ratio of $\epsilon_{Si}/\epsilon_{SiO_2}$. V_{dsat} is the potential at the channel “pinchoff” point. There are many models for V_{dsat} , among which the simplest one is $V_{dsat} = V_{gs} - V_{th}$, where V_{gs} is gate-to-source voltage and V_{th} is the threshold voltage. For short channel devices, V_{dsat} is channel length (L) dependent, and the relation is often modeled as [109]:

$$V_{dsat} = \frac{(V_{gs} - V_{th})LE_{cr}}{V_{gs} - V_{th} + LE_{cr}}, \quad (4.6)$$

where E_{cr} is the critical field for velocity saturation and its value is about $5 \times 10^4 \text{V/cm}$.

To monitor device degradation with easily measurable I_{sub} , the correlation between I_g and I_{sub} can be derived from Eq. 4.1 and 4.2.

$$\frac{I_g}{I_d} = C_2 \left[\frac{I_{sub}}{C_1 I_d} \right]^m, \quad (4.7)$$

where $m = \phi_b/\phi_i \approx 3$. This equation, which applies to the case where V_g exceeds V_d , has been verified in n-channel transistors. The NMOSFET device lifetime τ can also be modeled as:

$$\frac{\tau I_d}{W} \propto \left[\frac{I_{sub}}{I_d} \right]^{-\phi_{it}/\phi_i}. \quad (4.8)$$

This can be further simplified as

$$\tau \propto I_{sub}^{-m}. \quad (4.9)$$

For PMOSFET, the gate current is the determining parameter used in modeling [113, 114] and the lifetime model

$$\tau \propto I_g^{-m}. \quad (4.10)$$

The LEM has two major limitations [115]: (a) it relates HCD to the local field, thus neglecting the space and time tag of carriers in reaching local equilibrium with the field; (b) since the potential energy is the only source of energy available to the carriers, the maximum attainable energy is limited to qV_{tot} where V_{tot} is the total voltage drop experienced by the carriers. Therefore, the LEM predicts no HCD at voltages smaller than the threshold energy.

There are higher order models which attempt to overcome above limitations: (a) “non-local” LEMs; (b) effective temperature models (ETM). Non-local LEMs

replace the local field with “non-local” quantities such as the potential drop along the current flowlines [116] or suitable electric fields [110]. ETMs assume quasi-equilibrium Maxwellian distributions whose effective temperature (T_e) is a function of the local field. T_e can be obtained from the energy conservation equation [117]. In their simplest form, LEM and ETM predict the relationships between I_{sub} and I_g :

$$I_{sub} \propto I_d \exp(-\phi_i/E^*), \quad (4.11)$$

$$I_g \propto B(E_{ox})I_d \exp(-\phi_b/E^*), \quad (4.12)$$

$$I_g/I_d \propto (I_{sub}/I_d)^{\phi_b/\phi_i}, \quad (4.13)$$

where $B(E_{ox})$ models the collecting efficiency of the gate; $E^* = q\lambda E_l$ for LEM, E_l is local electric field; $E^* = k_B T_e$ for ETM.

Although the higher order models and complex 2-D and Monte-Carlo simulation may provide more precise result, it is difficult to match precisely to the hardware, due to the complex physics involved the uncertainties in the doping profile and the device operating conditions. The I_g/I_{sub} model 4.7 is widely applied in reliability characterization for its simplicity and easy measurement.

4.3.2 Empirical Lifetime Models

4.3.2.1 Power Law Model

This model was proposed by Takeda [118] based on the following assumptions:

1. Avalanche hot carrier injection due to impact ionization at the drain, rather

than channel hot electron injection composed of “lucky electrons,” imposes the severest constraints on device design.

2. Device degradation (V_{th} shift and G_m change) resulting from drain avalanche hot carrier injection has a strong correlation with impact ionization induced substrate current.

The time dependence of any degradation parameter such as V_{th} shift, ΔV_{th} , or G_m degradation, $\Delta G_m/G_{m0}$, can be empirically expressed as

$$\Delta V_{th}(\text{or } \Delta G_m/G_{m0}) = At^n. \quad (4.14)$$

This expression is particularly valid for short stress times, while for long stress time, ΔV_{th} and/or $\Delta G_m/G_{m0}$ begins to saturate. The slope n , which relates to the interface trap generation, is strongly dependent on V_g but has little dependence on V_d . This suggests that n may change according to hot carrier injection mechanism. n was 0.5-0.6 ($L_{eff} = 0.35 - 2\mu m$ and $t_{ox} = 6.8 - 20nm$) [118] and 0.65 [109]. Hu et al. demonstrated that n is expected to be between 0.5 (diffusion limited) and 1 (reaction limited). The magnitude of degradation, A , is strongly dependent on V_d and has little dependence on V_g . In particular,

$$A \propto \exp(-\alpha/V_d), \quad (4.15)$$

Therefore, the lifetime τ can be expressed as

$$\tau \propto \exp(b/V_d), \quad (4.16)$$

where $b = \alpha/n$.

Takeda [99] and Hu [109] both reported $\tau \propto I_{sub}^m$, while m ranging between 3.2-3.4 given by Takeda and 2.9 by Hu.

4.3.3 Dynamic Stress Lifetime Modeling

It is essential to understand the device hot carrier degradation under dynamic stress in the real operating circuit. Many experiments and analyses on AC hot carrier effects have been done. Takeda et al. [119] proposed a guideline to model the AC hot carrier effect with precaution against noise: (1) AC HCD in LDD and GOLD structures can be estimated on the basis of DC degradation in terms of effective stress time, which takes the duty factors into account, (2) for LDD structures, no degradation specific to AC stress exists for frequency up to at least 10MHz. Quader et al. [120] proposed a generalized DC to AC lifetime conversion factor by considering the operation frequency, input rise and fall times. The NMOSFET lifetime factor is

$$NTF = \frac{4}{ft_r}, \quad (4.17)$$

and the PMOSFET lifetime factor is

$$PTF = \frac{10}{ft_f} \quad (4.18)$$

where f is the operating frequency and t_R, t_f are the gate signal rise and fall times. For bi-directional circuits in which both the gate and the source voltage can change during transients, the first order NTF and PTF were also proposed.

$$NTF \approx \left(\frac{1}{f}\right) \frac{1}{time[V_{ds} > 0.93V_d; 0.1V_d < V_{gs} - V_{tn} < 0.35V_d]} \quad (4.19)$$

$$PTF \approx \left(\frac{1}{f}\right) \frac{1}{time[V_{sd} > 0.93V_d; 0 < V_{sg} - V_{tp} < 0.1V_d]} \quad (4.20)$$

4.3.4 Temperature Dependence Model

The temperature dependence of HCD can be modeled by the Arrhenius relationship. The activation energies was reported around $-0.1eV \sim -0.2eV$ [41].

$$\tau \propto \exp\left(\frac{E_a}{kT}\right). \quad (4.21)$$

The negative activation energies were from early reports which showed that hot carrier effects are enhanced at low temperature. The main reason for this is an increase in the electron mean free path and the impact ionization rate at low temperature. As shown in [99], substrate current at 77K is five times greater than that at room temperature (RT) and CHE gate current is about 1.5 orders of magnitude greater than that at RT. The effects of a given oxide degradation are increased at low temperature [121], [122], [9] probably because electrons with low thermal energy have difficulties overcoming the potential barrier. These cause the worse degradation of performance at low temperature than that at RT. For NMOSFET, the CHE mode causes more severe degradation than the DAHC at low temperature because of an increase in electron trapping and a reduction of interface state generation, while at RT the CHE mode is less severe. Hu [109] showed the temperature coefficient of CHE gate and substrate current to be negative.

4.3.5 Statistical Model

Generally, lognormal distribution is utilized to model HCD lifetime. Snyder et al. [123] did HCD lifetime test of total more than 1,000 NMOS transistors (from two companies) to demonstrate the goodness-of-fit of several lifetime distributions. Lognormal distribution was showed to have a better fit than normal distribution and one extreme value distribution. Kim and Hwang [124] discussed hot carrier lifetime variation caused by non-uniformity of the gate length. Kuntman et al. [125] applied Weibull distribution to model the threshold voltage degradation caused by hot carrier injection.

Chapter 5

Electromigration

5.1 Introduction

The dominating failure mechanism of interconnects, electromigration, is characterized by the migration of metal atoms in a conductor through which high current densities pass [4]. It is generally accepted that electrons streaming towards the anode can impart sufficient momentum to atomic ions upon impact to propel them into neighboring vacant sites. Many detailed aspects of EM are still under investigation although it has been studied for more than 40 years. The reason for this is the existence of many factors that influence EM and the inability to isolate the effect of these factors experimentally. Some of these factors are related to the interconnect materials and manufacture process, such as grain structure, grain texture, interface structure, film composition, physics of void nucleation and growth [126]. Current density and temperature are the two most important stress factors in modeling interconnect EM lifetime and design of accelerated testing. The well-known Black's equation of EM lifetime has an inverse power law relation with the current density and an Arrhenius relation with the interconnect temperature [127].

As expected, aggressive interconnect scaling has resulted in increasing current densities and associated thermal effects that will cause reliability problems. This will greatly reduce the interconnect lifetime if not well handled. Interconnects are now a

significant limiter and are as important as transistors in determining an IC's density, performance and reliability. Aluminum, the once major on-chip interconnects, has been gradually replaced by copper because copper has lower resistivity and higher electromigration resistance.

In order to understand and improve interconnect reliability performance, physical models and statistical models must be carefully built. In this chapter, EM physical process, lifetime and statistical models are briefly discussed.

5.2 Physics of Failure

EM has been the subject of intense study since Al thin film conductors were found to exhibit EM. A detailed review of past and recent studies can be found in [126, 128, 129, 130] and [131]. As device density increases, the interconnect that carries signals are consequently reduced in size – in height and cross-section. This leads to extremely high current densities, on the order of at least $10^6 A/cm^2$. At these current densities, momentum transfer between electrons and metal atoms becomes important. The transfer, which is called the electron-wind force, results in a mass transport along the direction of electron movement. Once the metal atoms are activated by the electron wind, they are subject also to the electric fields that drive the current. Since the metal atoms are positively ionized, the electric field moves them against the electron wind once they have been activated. The interplay of these two phenomena determines the direction of net mass transfer. This mass transfer manifests itself in the movement of vacancies and interstitials.

The vacancies coalesce into voids or microcracks, and interstitials become hillocks. The voids, in turn, decrease the cross sectional area of the circuit metallization and increase the local resistance and current density at that point in the metallization. Both the increase in local current density and in temperature increase EM effects. This positive feedback cycle can eventually lead to thermal runaway and catastrophic failure. Figure 5.1 summarizes the general EM failure process.

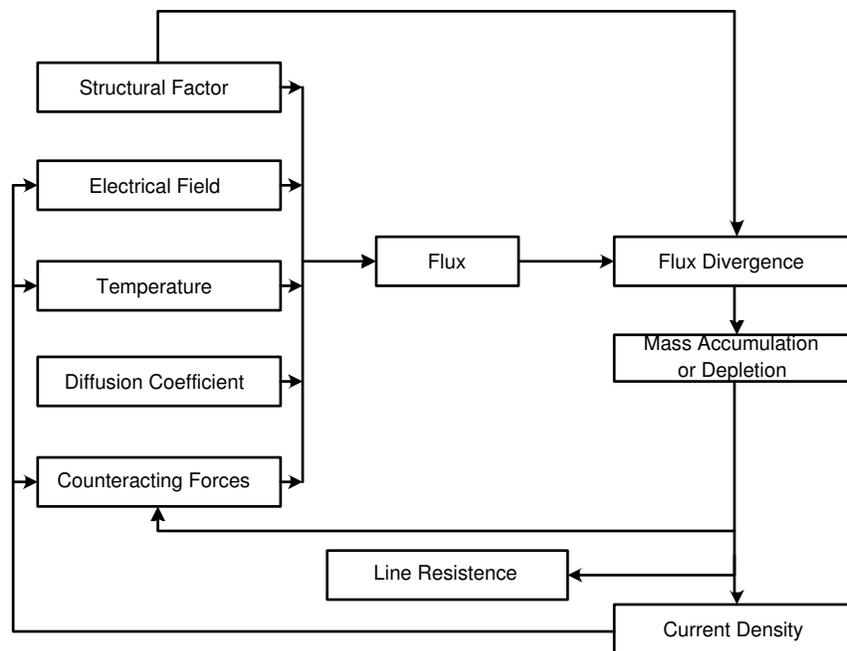


Figure 5.1: Summary of EM failure process [10].

Peckerar [132] illustrated the electron-defect interaction process. The fundamental process is reiterated below. Mechanical defects and grain boundaries in the interconnect play significant roles in void and hillock formation. Once an electron is scattered by an ion in a defect site, the electron's momentum is reversed. This causes an average change in momentum in the transport direction equal to $2m\bar{v}$ where m

is the electron mass and \bar{v} is the mean velocity of the electron in the direction of the current flow. The friction force on the ion is:

$$F_f = \frac{2m\bar{v}}{\tau_{col}}, \quad (5.1)$$

where τ_{col} is a time constant representing the time between collisions. From elementary transport theory, the electron current density J_e is:

$$J_e = ne\bar{v}, \quad (5.2)$$

where n is the density of electrons available for transport. From Eq. (5.2), the \bar{v} can be found and the friction force can be expressed as:

$$F_f = \frac{2mJ_e}{ne\tau_{col}}. \quad (5.3)$$

Assuming the ion transport is proportional to the force applied:

$$v_i = \mu F_f, \quad (5.4)$$

and

$$J_i^f = eN\mu F_f, \quad (5.5)$$

where v_i is the mean ion velocity in the direction of transport, μ is the ion mobility, J_i^f is the ion current density due to electron momentum transfer, and N is the density of ions available for transport. According to the Nernst-Einstein equation, the ion diffusion coefficient, D , and μ are related as:

$$\frac{D}{\mu} = \frac{kT}{e}, \quad (5.6)$$

where k is the Boltzmann constant. Combining Equations (5.4)-(5.6) yields:

$$J_i^f = \frac{e^2DN}{kT} F_f. \quad (5.7)$$

And from Eq. (5.3), F_f is proportional to J_e :

$$F_f = C_1 J_e, \quad (5.8)$$

where C_1 is the proportionality constant derived from Eq. (5.3). Thus we get:

$$J_i^f = \frac{e^2 D N C_1}{kT} J_e. \quad (5.9)$$

From Ohm's law, this equation becomes:

$$J_i^f = \frac{e^2 D N C_1}{kT} \rho_e E, \quad (5.10)$$

where ρ_e is the electron resistivity and E is the electric field.

The electric field will also induce an ion current, J_i^E , which is counter to the friction current. Using the basic transport relation and Nernst-Einstein relationship to describe the field-induced ion current J_i^E :

$$J_i^E = \frac{e^2 N D E}{kT}. \quad (5.11)$$

The total ion current, J_i , is given by:

$$J_i = J_i^f - J_i^E = (eN)(C_1 \rho_e - 1) \left(\frac{eD}{kT} \right) E. \quad (5.12)$$

The simple interpretation of this equation is that the ion current is equal to the effective charge on the ion, multiplied by the density of ions available for transport, the ion mobility, and the electric field.

There are other physical effects that may give rise to net ion currents and to the ion current divergence necessary for void formation. The temperature gradients occurring in the interconnect will create the ion flux divergences responsible for

open-metal device failures because the ion diffusion coefficients will become position dependent. Mobilities will be greater in the hotter region and less in the cooler region. The ion will come from the hotter regions to the cooler regions to form hillock.

The stress in the conducting strip also affects the EM [133]. Just as an electron field causes ion drift, a gradient of stress σ acts as a generalized force to induce ion motion. Ions preferentially migrate from compressively (σ more negative) stressed regions and accumulate at locations stressed in tension (σ more positive), while vacancies diffuse the other way. The resulting stress gradient causes a backflow of matter, this effect plays a significant role in short conductors. Furthermore, compressive and tensile stresses may increase or decrease ion migration activation energy, changing the diffusion coefficient D .

Scherge et al. [134] proposed a more complete equation that accounts for these effects is:

$$J_i = e(C_1\rho_e - 1)N\left(\frac{eD}{kT}\right)E - eD\nabla N - \frac{eDN\omega\nabla N}{\beta N_0 kT}, \quad (5.13)$$

where ω is the atomic volume, β is the film compressibility, and N_0 is the atom density of the film. The first term accounts for friction flow, the second accounts for the concentration gradient, and the third accounts for film stress.

The electrically-induced diffusion of metal atoms alone is not sufficient to cause EM. The growth of voids and hillocks requires the saturation of vacancies and the supersaturation of interstitials [10], which requires not only a diffusion of metal atoms but also a divergence in the diffusion flux.

Microscopic examination of most deposited thin films indicates a pronounced cellular structure to the film generally referred to as the “grain structure” of the film. These cells arise as a result of the processes of nucleation and growth that form the film. The grain structure depends on the deposition conditions and has a profound effect on EM damage. For example, powered single-crystal Al strips have been shown to exhibit virtually “infinite” life.

Grain density is determined by surface conditions and film growth parameters, such as substrate temperature, rate of arrival of metal atoms to the growth surface. Grain boundaries represent interfaces with associated free energy of surface formation. During growth and subsequent annealing cycles, some grains may grow and others disappear in order to minimize the free energy. The grain boundaries represent relatively low-resistance ion conducting channels. At standard IC operating temperatures, bulk ion migration processes are slow and the grain boundaries carry the bulk of the ion current [132]. The grain structures and boundaries enable considerable refinement of material models of EM. Specifically, there are three properties that have immediate impact on reliability models. They are:

- The orientation of the boundary with respect to the electric field.
- The angles of the grain boundaries with respect to each other.
- Changes in the number of the grains per unit area—grain density.

Each of these properties can give rise to the ion divergences necessary to create voids in metal strips. The effects of these properties will be discussed below.

Broad area or blanket metallization leave greater numbers of grain boundary “triple points” and grains lineup in bamboo structures after patterning when the interconnect stripe widths shrink. Figure 5.2 shows the confluence of three grain boundaries at a triple point. If the boundary to the left is parallel to the applied field, the angle θ_1 equals to 0, the apparent ion mobility is highest along that boundary. Migration along the two adjacent boundaries is the result of a projected field component and is lower. Under this condition, it is apparent fewer ions leave the triple point than enter it, and a mass accumulation is favored. Otherwise voids form.

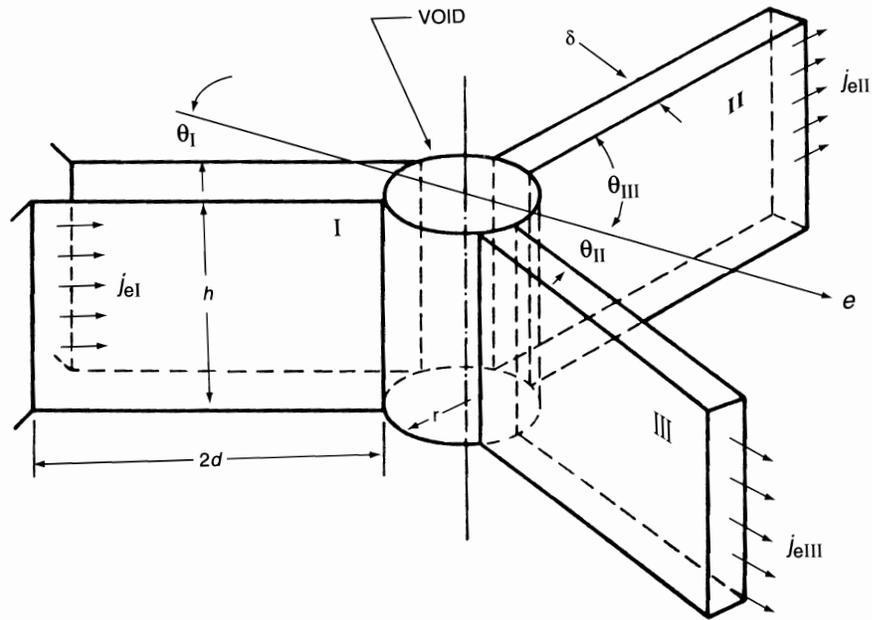


Figure 5.2: Schematic of grain-boundary triple point configuration [4].

If the grain-size changes along the strip, the density of ion conduits into and out of a region must also change. A densely grained region will channel ions out more effectively than a sparsely grained region. This creates the ion current divergence

necessary to form a void. Similarly, ion pile-up can occur in regions in which the grain sizes increase in the direction of electron flow. Another factor that affects the ion conduction is the texture of the oriented crystallite in the metal films.

To summarize, the important factors of EM include current density, electrical field, temperature, grain structure and boundary. An accurate model should consider all these factors. This will definitely increase the complexity of the model and the need to conduct a bunch of well-designed experiments to build the model. Under this condition, an empirical lifetime model and the statistical lifetime distribution can assist engineers to understand the EM mechanisms better and realize reliability design goals.

5.3 EM Models

5.3.1 Lifetime Models

5.3.1.1 Black's Equation

Black [127] developed an empirical model relating the median time to failure (t_{50}) of a metal film conductor to the temperature (T) and current density (J).

$$t_{50} = \frac{A}{J^2} \exp\left(\frac{E_a}{kT}\right), \quad (5.14)$$

where A is a material and process dependent constant, E_a is the activation energy for the diffusion processes that dominate over the temperature range of interest, k is the Boltzmann's constant and T is temperature in Kelvins. The usage of 2 as the current exponent has raised questions since it has been argued that the mass

flux divergences responsible for failure are proportional to the current density (the exponent is expected to be unity). Considering vacancy accumulation due to a divergence in the EM-induced vacancy flux and normal Fickian diffusion, Shatzkes and Lloyd [135] theoretically derived that current density has an exponent of 2. Lloyd [136] further proposed a generalized Black equation:

$$t_{50} = A \cdot J^{-n} \cdot T^{-m} \cdot \exp\left(\frac{E_a}{kT}\right). \quad (5.15)$$

The various values of n and m are determined by the particular failure physics and conductor's geometry. If $n = 2, m = 0$, it is the original Black model. In respect to failure physics, for all nucleation-dominated failures, $n = 2, m = 0$; if a failure is growth-dominated, $n = 1, m = 0$. In respect to conductor's geometry, it has been observed that for wide lines, defined as those where the average grain size is smaller than the line width, $n = 2$, whereas for narrow lines, $n = 1$. For engineering applications, it proves that there is no significant difference between which n and m values are used, however, calculations showed the combination of $n = 2$ and $m = 2$ would produce very good lifetime predictions and the extrapolated activation energies being reasonably accurate [136]. The side effect of using the generalized Black equation is the nonlinearity when extrapolating activation energy, which makes parameter extraction very hard.

Thermal activation energy E_a can be extracted from accelerated lifetime tests. It typically reflects a complex admixture of mass transport and damage processes: grain-boundary diffusion, surface diffusion, stress-assisted diffusion, void nucleation, extrusion and so on. A detailed list of activation energies of Al-based interconnect

(before 1991) can be found in [128]. Ohring [4] summarized activation energy data of Al and Al alloy:

1. The activation energy for bulk diffusion in pure Al is 1.4 eV. For films with large grain size, E_a is found to be 1 to 1.2 eV. E_a of usual fine-grained films ranges from 0.4 to 0.6 eV which indicates grain-boundary mass transport-induced damage.
2. Additions of Cu to Al raise E_a by a few tenths of an eV and extend mean-time-to-failure (MTTF) lifetimes accordingly. This is the basis for the use of Al alloy metallizations containing 0.3 to 5 wt.% Cu. It is also common to add 1 wt. % Si as well to prevent Al-Si inter-diffusion at semiconductor contacts.
3. E_a scales directly with higher melting point and larger grain size.

For copper interconnect, the fast diffusion paths were found to be at interfaces and varies depending on the fabrication process and materials. A range of activation energies for copper have been reported from 0.7 to 2.4 eV [137]. Lists of copper activation energies can be found in [138] and [139].

5.3.1.2 EM Lifetime Dependence On Interconnect Width and Length

EM lifetime is a function of the linear dimensions of the interconnect stripe. Since the EM-induced open-circuit failure must occur along the interconnect width, an increase of lifetime with increasing line width (exceeding the metal grain size) is expected because the probability of aligning defects to make a wider interconnect

open is smaller. However, if the interconnect width is reduced to comparable to or smaller than the mean grain size, the lifetime is found to either level off or increase because the interconnect has a ‘bamboo’ structure and the mass transport of the metal ions is more difficult to occur as there are less triple points in the line.

EM lifetime dependence on interconnect length has been investigated in several papers [140, 141, 142]. Series model [142] was applied to model the length effect by treating the interconnect as a series connection of multiple unit lines. The series model predicts that the EM lifetime approaches zero as the interconnect is long enough. However, empirical evidence [140, 141] has shown that the failure time decreases rapidly with increasing line length and then reaches a saturation value beyond a critical length. These experimental results reveal that EM lifetime is dependent on the most severe defect in the whole interconnect, rather than on the number of the defects with severity beyond a certain level [143]. The length increase will increase the possibility of finding more severe defects, but the severity of extrinsic and intrinsic defects has an upper bound because of the process stability and material structure.

On the basis of the above results, the EM lifetime dependence on the interconnect length and width may have an empirical form [144]:

$$t_{50} \propto AW\Theta \exp\left(\frac{\alpha}{L}\right), \quad (5.16)$$

where Θ , W , and L is the interconnect thickness, width and length, respectively. A is a constant, α is a constant dependent on width and thickness.

5.3.1.3 EM Lifetime Dependence On Current Waveforms

Interconnect experiences unidirectional or bidirectional AC current stressing in circuit environment. Based on the DC current stress model, many models have been proposed to model the EM lifetime under AC current stress.

For unidirectional pulsed current, the average current model has been proposed by Towner [145] and Brook [146]. Other models, including the vacancy supersaturation model by Clement [147] and the defect relaxation model by Tao et al. [148], have been proposed to model the pulsed current stress. These models all validated the average current model from different points of view. The validity of the average current model under low and high frequency has been verified by Liew [149] and Tao [148].

The average current density J_{avg} is defined as

$$J_{avg} = \frac{1}{t_p} \int_0^{t_p} J(t) dt \quad (5.17)$$

where t_p is the period of the current pulse. And the t_{50} is

$$t_{50}^{pulse} = A J_{avg}^{-2} \exp\left(\frac{E_a}{kT}\right). \quad (5.18)$$

For bidirectional current stress, it has been observed that the EM damage incurred by the forward current stress can be partially healed by the following reverse current. Ting et al. [150] proposed an average current recovery model

$$t_{50}^{Bi-di} = A J_{eff}^2 \exp\left(\frac{E_a}{kT}\right). \quad (5.19)$$

The effective current density J_{eff} is defined as

$$J_{eff} = |\bar{J}^+| - \gamma |\bar{J}^-| \quad (5.20)$$

where \bar{J}^+ and \bar{J}^- are average current density including only the positive and the negative currents, respectively. γ represents the degree of damage recovery due to the current with opposite polarity. It has a range from 0 to 1, $\gamma = 0$ means no healing effect, whereas $\gamma = 1$ means perfect healing. Usually γ is around 0.9.

5.3.1.4 EM Lifetime Dependence On Joule Heating Effect

The exponential dependence of EM lifetime on temperature makes the interconnect temperature an important factor in modeling EM degradation, especially in EM characterizations test, which features high stress temperature (around 300 °C) and high current density (25 mA/ μm^2). Joule heating caused by the high current density increases the stress temperature and generates temperature gradients in the conductor line that promotes EM activity [151].

With the introduction of copper interconnect and low-k inter-dielectric material in advanced technologies, the Joule heating effect becomes more important because low-k materials generally have poorer thermal conducting properties than the traditional oxide [152], and the current density is increased because of the larger scale integration. In most cases of normal operation, the power distribution network is the most vulnerable part to EM in the chip. The local temperatures of these power buses should be considered in modeling the EM lifetime of the whole chip.

Many works have been done to include the Joule heating effect into EM lifetime modeling [153, 154, 155, 156, 157]. To consider the Joule heating either from the metal line itself or from its neighboring lines, the actual metal stress temperature

can be modeled:

$$T_m = T_s + \Delta T, \quad (5.21)$$

where T_s is the substrate temperature. The temperature increase due to Joule heating, ΔT , is determined by the line power generation and thermal resistance of the system [153].

$$\Delta T = \frac{J^2 \rho_0}{\frac{K_i}{\Theta t_i} [1 + 0.88 t_i / W] - J^2 \rho_0 \beta_m}, \quad (5.22)$$

where J is the current density, ρ_0 is the resistivity of the metallization, β_m is the temperature coefficient of the resistivity, W is the interconnect width, t_i is the insulator thickness, Θ is the metal thickness, and K_i is the thermal conductivity of the insulator. These parameters need to be determined according to the stacks and layouts of the metal interconnects and the surrounding dielectric materials [157].

And the Black's equation is then:

$$t_{50} = A J^{-n} \exp\left[\frac{E_a}{k(T_s + \Delta T)}\right] \quad (5.23)$$

5.3.2 Statistical Models

5.3.2.1 Lognormal Distribution

Lognormal failure distribution has been used to characterized EM lifetime. Assuming that the time to failure t is a random variable, the lognormal probability density function $f(t)$ is:

$$f(t) = \frac{1}{\sigma t \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{\ln(t) - \ln(t_{50})}{\sigma}\right)^2\right], \quad (5.24)$$

where t_{50} is the median time to failure and σ is the lognormal standard deviation. t_{50} can be estimated by the well-known Black's equation. For lognormal standard deviation σ , which may be related to the ratio of the linewidth to the grain size [158] and current density [159], ranges from 0.28 to 1.4 [160] from early research.

Various arguments have been presented to justify the use of the lognormal distribution.

Normally distributed activation energy. Schwarz [161] used temperature-ramp resistance analysis to determine the distribution of activation energies for EM damages in Al and Al-4%Cu thin film interconnects. He found that the activation energies for the pure aluminum conductors are well represented by a normal distribution. For Al-4%Cu interconnects, there are three normal distributed subpopulation of activation energies.

Normally distributed conductor temperatures. Lloyd [162] demonstrated that a normal temperature distribution, given the variation is small compared to the mean temperature, can produce a lognormal failure distribution in EM lifetime experiments. Bobbio et al. [163] also justified the applicability of the lognormal distribution based on the temperature dispersion of conductors during life testing.

Lognormal distributed grain sizes. Based on the grain boundary diameter distribution that was microscopically measured, Attardo et al. [164] used Monte Carlo simulation to get the failure times. Their simulation results determined that the statistical failure rate distribution best fits a lognormal curve for wide

lines.

Experiment results. Towner [165] performed EM lifetime tests on a variety of Al alloy films using sample sizes ranged from 35 to 120. His results showed that the lognormal distribution fit better than the logarithmic extreme distribution where the grain size is smaller than the linewidth. Where the grain size exceeds the linewidth, either distribution can be used to represent the data. Gall et al. [166] did an experiment utilizing large interconnect arrays in conjunction with Wheatstone Bridge. Over a temperature range from 155 to 200°C, a total of more than 75,000 interconnects were tested. The results indicated that the EM failure mechanism in this experiment follow perfect lognormal behavior down to the four sigma level.

Although lognormal distribution is widely used to model EM lifetimes, it cannot be used as an element failure distribution that can be applied with the “weakest link” model. If $F_1(s)$ is the cumulative distribution function (CDF) of strength of a single link, then statistically the CDF of a chain of N independent (in strength) links is:

$$F_N(s) = 1 - [1 - F_1(s)]^N. \quad (5.25)$$

In general, the form of the distribution F_N depends on the number of links (failure elements) in the chain (series), so “scaling up” the model for longer chains affects the choice of modeling distribution. The lognormal distribution does not scale in Eq. (5.25) and therefore cannot be the failure distribution for elements in series. This means lognormal distributions can approximate true failure times only in a

finite percentile interval.

5.3.2.2 Weibull Distribution

The Weibull probability density function can be expressed as:

$$f(t) = \left(\frac{\beta}{\alpha^\beta}\right)t^{\beta-1}\exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right], \quad (5.26)$$

where α is the characteristic lifetime and β is the shape parameter. Weibull distribution is “scaling,” and as the lognormal distribution, its physical meaning is not clear yet. Generally, the lognormal distribution fits the experiment data better than the Weibull distribution, but there is no significant improvement. Lifetest data of aluminum conductors showed that lognormal and Weibull distribution equally fit well at large percentage failures (0.1-1%), but at lower percentage failure, the projected failure rates differ by several order of magnitude [164]. There are different views about the usage of Weibull distribution. Gall et al. [166] used simulation results to roll out the Weibull distribution in the analysis of their experiment data. Pennetta [167] simulated EM damage in metallic interconnects by biased percolation of a random resistor network in the presence of degradation and recovery processes. Both the lognormal distribution and Weibull distributions fit the simulation result well.

5.3.2.3 Bimodal Distribution

Bimodal lifetime distribution is often seen in copper via EM tests. The copper via has been identified as a weak link. The majority of early failures can be

attributed to the copper vias [168]. Lai et al. [169] suggested two EM failure mechanisms—via related and metal-stripe-related. Ogawa et al. [170] reported two distinct failure modes in dual damascene Cu/oxide interconnects. One model was found to be a void formation within the dual-damascene via, the other was a voiding that occurs in the dual-damascene trench. Multiple failure modes have been reported in [171], [172], [173], [174], [175], [176],

A single lognormal distribution is observed only if one physical mechanism dominates the failure process. Failure distribution involving two different failure modes may appear nonlinearly in the lognormal probability plotting paper. Under this situation, bimodal lognormal distribution might be applied to model the data. Suppose there are two different EM failure mechanisms. Each individual mechanism is described by a lognormal distribution $CDF_A(t)$, (respectively $CDF_B(t)$) over time t , with median time to failure t_{50A} (respectively t_{50B}) and standard deviation σ_A (respectively σ_B). A priori the failure mechanisms may have different activation energies E_A and E_B as well as the current density exponents n_A and n_B . There are two different models of an overall bimodal failure distribution [177].

1. Superposition Model

Consider a sample in which the failure scenario is influenced by the presence or absence of a particular physical property in the test device. Its presence forces a specimen to fail due to mechanism A, its absence exclusively due to B. The property appears with a probability $P(A)$. The property is absent with

a probability $P(B) = 1 - P(A)$. The overall CDF of all the specimens is:

$$CDF(t) = P(A) \cdot CDF_A(t) + (1 - P(A)) \cdot CDF_B(t). \quad (5.27)$$

The resulting CDF appears *s*-shaped in the probability-plot.

2. Weak-Link Model

In this scenario different failure mechanisms can cause the interconnect failure in a serial fashion. If the failure mechanisms act statistically independent, the overall CDF is given by:

$$CDF(t) = 1 - (1 - CDF_A(t)) \cdot (1 - CDF_B(t)). \quad (5.28)$$

This CDF has a “hook-shaped” in the probability plot.

Chapter 6

Physics-of-Failure Based VLSI Circuits Reliability Prediction

Methodology

6.1 Introduction

With a history of more than five decades [178], reliability prediction has an important role in business decisions like the system design, parts selection, qualification, warranties and maintenance. Nowadays electronic system designers have their own industry-specific reliability prediction tools, such as the well-known MIL-HDBK-217, SAE reliability prediction method, Telcordia SR-332 and PRISM, etc [178, 179]. Many of those methods are empirically-based which were built upon field data and extrapolations such as “the parts count method” and “the parts stress method” with various kinds of pre-factors [24]. One big disadvantage of those empirical-based methods is the lack of integration of the physics-of-failure models because of the complexity and difficulty for the system designers to get detailed technology and microcircuit data [178]. Prediction accuracy is diminished without those PoF models, and the situation is becoming worse with technology advancement. Today’s microelectronic devices, featuring ultrathin gate oxide and very short channels, suffer various detrimental failure mechanisms, including TDDB [6], NBTI [180], HCD [181] and EM, as the non-ideal voltage scaling brings higher field and

current density. Each failure mechanism has its unique dependence on voltage and temperature stresses and all can cause device failure. The traditional prediction method is not applicable as before, considering the multiple failure mechanisms' effect and the difficulty to obtain enough up-to-date field data.

Device manufactures face the same challenge to maintain and further improve reliability performance of advanced microelectronic devices, in spite of all kinds of difficulties from technology development, system design and mass production. Conventional product reliability assurance methods, such as burn-in, HTOL and HALT, are gradually losing competitiveness in cost and time because the gap between normal operating and accelerated test conditions is continuing to narrow and the increased device complexity makes sufficient fault coverage tests more expensive [22]. An accurate reliability simulation and prediction tool is greatly needed to guide the manufacturers to design and deploy efficient qualification procedure according to customer's need, and help the designers to get intime reliability feedback to improve the design and guarantee the reliability at the very first stage.

The needs of accurate reliability prediction from both device and system manufacturers require integration of PoF models and statistical models into a comprehensive product reliability prediction tool that takes the device and application details into account. A new physics-of-failure based statistical reliability prediction methodology is proposed in this chapter. The new methodology considers the needs of both the device manufacturer and the system supplier by taking application and design into account. Based on circuit level operation-oriented physics-of-failure analysis, this methodology provides an application-specific reliability prediction which

can be used to guide qualification and system design.

6.2 Review of Reliability Prediction Tools

Denson briefly reviewed the development of reliability prediction since World War II [178]. In 1962, the first version of MIL-HDBK-217 was published by the U.S. Navy and quickly became the operational standard. Since then, many industry-specific reliability prediction tools have been developed to meet the segment's unique need. A non-exhaustive list of these tools is given below: SAE reliability prediction method, Telcordia SR-332 [182], CNET RDF-2000 [183], British Telecom HRD-5 [184], Siemens SN29500 [185], NTT procedure [186], PRISM [187] and FIDES [188].

Although the MIL-HDBK-217 series (final version MIL-HDBK-217F [24]) was discontinued in 1994, the prediction methodology has been adapted by many later tools. The handbook method assumes that component follows constant failure rate distribution. This brings simplicity and extendability to the prediction method but also draws criticism [189]. Validity of the assumption was questioned because field data may contain various kinds of failures which include infant mortality, wearout, overstress and human error.

The parts stress model and the parts count model are two basic methods in the handbook approach. The parts stress model is applied at component level to obtain part failure rate (λ_P) estimation with stress analysis. A typical part failure rate can be estimated as:

$$\lambda_P = \lambda_b \cdot \pi_Q \cdot \pi_E \cdot \pi_A \cdot \pi_T \cdot \pi_V, \quad (6.1)$$

where λ_b is the base failure rate obtained from statistical analysis of empirical data, the adjustment factors include: π_T (temperature factor), π_A (application factor), π_V (voltage stress factor), π_Q (quality factor) and π_E (environmental factor). The equipment failure rate (λ_{EQUIP}) can be further predicted through part count method.

$$\lambda_{EQUIP} = \sum_{i=1}^n N_i (\lambda_g \cdot \pi_Q)_i, \quad (6.2)$$

where λ_g is the generic failure rate for the i^{th} generic part, π_Q is the quality factor of the i^{th} generic part, N_i is the quantity of i^{th} generic part and n is the number of different generic part categories in the equipment.

To accommodate the advancement of technology, a reliability growth model was introduced in the handbook approach to reflect the state-of-the-art technology.

$$\lambda_P \propto \exp[G_r(t_2 - t_1)] \quad (6.3)$$

where G_r is the growth rate, t_1 is the year of manufacture for which a failure rate is estimated, t_2 is the year of manufacture of parts on which the data were collected. It takes time to collect field data and obtain the growth rate G_r , especially when the growth is fast. Furthermore, the validity of applying a reliability growth model without taking technology generation into consideration is not confirmed yet.

Prediction methods based on the handbook approach usually provide conservative failure rate estimation [190]. Many things can and should be done to improve the prediction accuracy.

1. Integration of physics-of-failure analysis and modeling. This is a prominent issue because advanced microelectronic devices are vulnerable to multiple failure mechanisms. These failure mechanisms have their unique voltage and

temperature dependence. No unified lifetime model can take all these into account.

2. Integration of failure mechanism lifetime distribution. CFR assumption might give a fast and cost-effective failure rate estimation at the system level because the inherent inaccuracy is so obvious without justification from detailed PoF analysis. Weibull distribution has been demonstrated as the best fit lifetime distribution for TDDB. Lognormal distribution has been experimentally verified for HCD and EM. These specific lifetime distributions should be utilized in prediction for better modeling component and system lifetime.

6.3 Methodology

The PoF based statistical approach models device reliability by considering all the intrinsic failure mechanisms under dynamic stresses. Generally speaking, today's microelectronic device integrates many functional blocks which consist of thousands even millions of transistors. Running a full spectrum simulation will consume unacceptable amounts of resources. To reduce the simulation complexity and release the heavy load of computation, the proposed PoF statistical reliability prediction methodology takes four unique approaches by considering the repetitive characteristic of CMOS circuits.

Cell-Based Reliability Characterization

Standard cells (inverter, NOR, NAND, etc) are the fundamental building blocks in modern VLSI circuit design. Cells in same category have similar

structures and operation profiles. For instance, the SRAM chip consists of millions of bit cells that have the same transistor configuration. These SRAM bit cells have similar operations applied to them: read, write or hold. For chips as complex as a microprocessor, it can still be divided into functional block and further the cells.

In the PoF statistical approach, reliability characterization starts from the standard cells. Doing this can ease system designers' concern of understanding circuit details and running circuit simulation. Advantages of cell-based reliability characterization are listed below.

- Time saving in circuit design. Cell schematic and layout can be obtained from design kits provided by semiconductor manufacturers or design houses. There is no need for system designers to understand the circuits from the very beginning since they often are not electrical engineers. What they need to understand is the categorization of the cells and the reliability character of each category.
- Time saving in circuit simulation. The VLSI device simulation requires detailed circuit information and consumes lots of computation resource. Since the goal of circuit simulation is to find out the stress profile of transistors, cell level simulation provides a better option because of the small count of transistors inside a cell.

Equivalent Stress Factor

Equivalent stress factor (ESF) is used to convert dynamic stresses to static

stresses that have the same degradation effect. For each failure mechanism, lifetime model is built upon acceleration tests, which are generally carried out with highly accelerated static voltage and temperature stresses. However, a transistor in real operation has a dynamic stress profile, and the static PoF models can't be applied directly.

The ESFs are obtained through cell reliability characterization and then applied in device reliability prediction. These factors are specified to cell, transistor, operation and failure mechanism. Given a specific cell, for each operation, the voltage and current stresses of each transistor in the cell are obtained through SPICE simulation. For each failure mechanism, degradation under different stress conditions is accumulated and converted to an equivalent-total-stress-time (ETST) under a specified static stress condition by utilizing appropriate acceleration models. The ESF is the ratio of the ETST to the real stress time. To estimate cell reliability in a real application, the cell operation profile needs to be determined at first. The next step is using ESFs to calculate the “effective” stress time for each failure mechanism of each transistor. The cell reliability is estimated as a series system in which each transistor inside the cell corresponds to a component.

Best-Fit Lifetime Distribution

To improve the prediction accuracy, the best-fit lifetime distribution for each failure mechanism should be taken instead of using the CFR model without justification. In the PoF statistical approach, Weibull distribution is used to

model TDDB failures, and lognormal distribution for NBTI, HCD and EM. The cell is considered as a series system with each component corresponding to one failure mechanism.

Time-Saving Chip-Level Interconnects EM Analysis

Chip-level EM analysis becomes more important as IC complexity is always driven up by scaling. Although EM becomes more serious in submicron designs, it is limited to the power distribution network in most cases [191]. Circuit designers must follow design rules, which set the interconnect dimension and current limits. Many industry tools have been developed to help circuit designers check the EM hotspot, such as *VoltageStormTM* from Cadence Design Systems, and *RailMillTM* from Synopsys. To optimize the EM resistance, lower level interconnects are designed to be EM-failure-free by considering the Blech effect [192]. The power network becomes the weakest link because of the large current density it carries and the local Joule heating effect. This has been verified by acceleration test results [193].

In the PoF statistical approach, chip-level EM analysis is focused on the power network since all designs should pass the design rules check, and final products must survive the high-temperature, high-voltage defect screening. This provides a good approximation without running full-detailed interconnect network EM analysis.

A flowchart of the PoF based statistical method is shown in Fig. 6.1. A detailed description of each step of the procedure is discussed below.

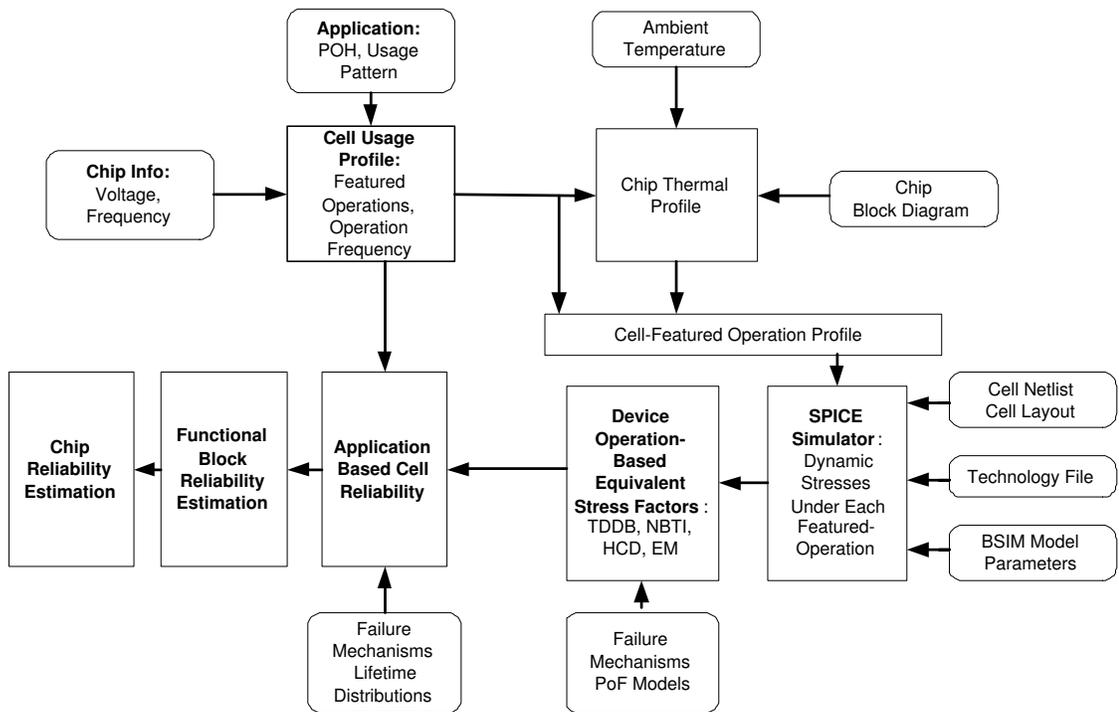


Figure 6.1: Flowchart of the physics-of-failure based statistical reliability prediction method.

6.3.1 Assumptions

The assumptions of the PoF based statistical approach are briefly explained below.

1. Degradation discretion and accumulation.

For each failure mechanism, the degradation process can be discretized by dividing the whole stress period into small intervals in order to accurately model the dynamic stress.

2. Negligible NBTI recovery effect.

NBTI degradation has been observed to have a recovery effect in acceleration tests after the stress has been removed [88, 194]. Physical understanding of this phenomena is still not clear. Since NBTI is a long term reliability concern, and recovery only has significant effect at early stages of stress [84], the NBTI degradation recovery is not considered in the PoF statistical approach.

3. Independent failure mechanisms.

All the failure mechanisms are assumed to be independent to each other. Each failure mechanism has its specific degradation region inside the transistor. TDDB causes damage inside the gate oxide while HCD/NBTI increase interface trap density. For PMOS, HCD and NBTI has been reported to be independent [195]. There is no confirmation about the interaction in field failure from literature research.

4. Competing failure modes.

The device is treated as a series system in which any cell failure will cause device failure. Every cell is viewed as a series system with each failure mode composing a block of the series system.

6.3.2 Input Data

In order to carry out reliability simulation and prediction, all the following information needs to be gathered: device application profile, device structure, cell schematic and layout, failure mechanism lifetime model and statistical distribution.

1. Application profile.

The device application profile can be broken down into operation phases with distinguishable environment factors. The PoF statistical approach deals with intrinsic failure mechanisms only, and the input data of each operation phase should include the ambient temperature (T_A) and the operating status-power on hours (PoH). Other factors such as humidity and vibration are not considered since they are mostly related to mechanical reliability.

2. Device structure and operation.

The PoF statistical approach takes a divide-and-conquer way to reduce the complexity of reliability simulation of VLSI devices. A device functional diagram is needed to divide the whole chip into functional blocks. Inside each functional block, cells are categorized and analyzed. Device operation needs to be analyzed to build cell operation profile.

3. Cell reliability simulation inputs.

- (a) Cell schematic and layout.
- (b) Technology file. This should be obtained from device manufacturers.
- (c) SPICE models.
- (d) Stimuli file. The stimuli file should be application-oriented so that the reliability output can be directly correlated to the stresses in real application.

4. Failure mechanism models and parameters.

- (a) Lifetime models and parameters.

With the technology information (t_{ox} , V_d , etc.) and acceleration test data, reliability engineers can choose or build the appropriate lifetime model for each failure mechanism. Lifetime models based on PoF analysis have been briefly reviewed in Chapter 2, 3, 4 and 5. Once the lifetime models have been decided, the model parameters can be estimated from maximum likelihood estimation (MLE) analysis or other regression analysis of acceleration test data.

- (b) Failure distributions and parameters.

It is important to have the correct failure distributions to estimate device reliability. For EM, lognormal distribution is normally the first choice. Weibull distribution has been widely used to model TDDB failures. For HCD and NBTI, lognormal distribution can be utilized. With given acceleration test data, the goodness-of-fit of these statistical distribution can be checked and the related distribution parameters can be estimated.

6.3.3 Device Thermal Analysis

Power dissipation of modern microelectronic devices has been rapidly increasing along with increasing transistor counts, clock frequencies and subthreshold leakage currents. The maximum power consumption of Intel microprocessors has been observed to increase by a factor of a little more than 2X every four years [196]. Detailed device thermal analysis becomes more important since most of the failure mechanisms are thermally activated. TDDB, NBTI and EM all have their own positive activation energy, only HCD has a negative activation energy which means HCD degrades faster at lower temperatures. For VLSI devices, thermal analysis should be carried out at functional block level because each block may have its own application pattern and the temperature across the whole chip might not be uniform. Several tools have been developed to do detailed chip level thermal-electrical analysis, such as ILLIADS [143] and HotSpot [197]. However, running this kind of tool requires very detailed circuit information and takes time for a system engineer to understand and learn. In the PoF statistical approach, device thermal analysis is carried out at the functional block level in order to get a quick and reasonable temperature estimation. Each functional block is assumed to have a uniform temperature. With a given application profile, the average substrate temperature of a block can be estimated as:

$$T_S = T_A + P_{total} \cdot R_{sa}, \quad (6.4)$$

where P_{total} is the total power dissipation of the block, and R_{sa} is the substrate-to-ambient thermal resistance. R_{sa} can be approximated by [143]

$$R_{sa} = \frac{1}{10^5 \cdot DieSize} \quad (6.5)$$

P_{total} is comprised of three main sources, dynamic switching power P_{dyn} , leakage power $P_{leakage}$ and switching power P_{sw} .

$$P_{total} = P_{dyn} + P_{leakage} + P_{sw} \quad (6.6)$$

P_{dyn} is due to charging and discharging of capacitive load which mainly consists of wiring capacitance [198].

$$P_{dyn} = a \cdot C_{total} \cdot V_d^2 \cdot f_c, \quad (6.7)$$

where a is the activity coefficient, V_d is the supply voltage, C_{total} is the total capacitive load of the wiring network which can be obtained through RC extraction, and f_c is the clock frequency.

There are two sources of leakage power: diode leakage power and subthreshold leakage power [143]. The diode leakage power is often neglected because it's generally small when compared with other power components. $P_{leakage}$ can then be expressed by

$$P_{leakage} = V_d \cdot I_{sl}, \quad (6.8)$$

where I_{sl} is the subthreshold leakage current, and

$$I_{sl} = K_0 \cdot \exp\left(\frac{V_{gs} - V_{th}}{nV_t}\right) \cdot \left(1 - \exp\left(-\frac{V_{ds}}{V_t}\right)\right), \quad (6.9)$$

where K_0 is a function of technology, V_t is the thermal voltage (kT/q), V_{th} is the threshold voltage and $n = 1 + \frac{\epsilon_{si}}{\epsilon_{ox}} \cdot \frac{t_{ox}}{D}$, where t_{ox} is the gate oxide thickness, D is

the channel depletion width, ϵ_{si} is the relative permeability of silicon and ϵ_{ox} is the relative permeability of oxide.

Switching power P_{sw} is a consequence of the gate input signal transition causing a charge or discharge of certain internal capacitances. A simple estimate is

$$P_{sw} = I_{sw} \cdot V_d, \quad (6.10)$$

where I_{sw} is the average switching current flow, which can be obtained through circuit simulation [199].

6.3.4 Cell Reliability Estimation

6.3.4.1 Dynamic Stress Profile Extraction

SPICE simulation can provide continuous output of these stress parameters like voltage and current. For reliability estimation involving multiple failure mechanisms, the simple average stress over time is not accurate enough because most of the failure models have exponential or power law dependence on voltage or temperature. By choosing an appropriate sampling interval T_I , the continuous stress profile can be discretized by sampling periodically. By assuming each operation takes one clock cycle T_A , for transistor M_m ($m=1, 2, \dots, M$) in the cell, the stress profile in j th period ($j=1, 2, \dots, \lceil T_A/T_I \rceil$) is: V_{gs}^{ij} , V_{ds}^{ij} and I_{ds}^{ij} etc. For interconnect W_w ($w=1, 2, \dots, W$), the stress current in j th period is I_w^{kj} , and the stress temperature is T^j . A sample script under Cadence Spectre simulation environment can be found in Appendix A.

6.3.4.2 Equivalent-Stress-Factor Evaluation

ESFs are specific to cell, transistor, operation and failure mechanism. In general, cell transistors' degradation are not the same because each transistor may have its unique stress profile. For an example, HCD is normally observed in NMOS transistors while NBTI is only observed in PMOS transistors.

Given a transistor in a cell, with the extracted operation-specific dynamic stress profile, failure mechanism ESF estimation takes the following steps.

To simplify the derivation, let's take a general transistor M_0 as an example. For M_0 , voltage stresses in the j th period are gate-to-source voltage V_{gs}^j and drain-to-source voltage V_{ds}^j , and the stress temperature T^j . A standard stress profile is set as voltage $V_{gs}^S = V_{ds}^S = V_D$, V_D is the supply voltage, and temperature T_S .

1. TDDB ESF

Assume the TDDB lifetime dependence on voltage is modeled by the exponential law with voltage coefficient γ_{TDDB} and dependence on temperature is model by Arrhenius relationship with activation energy E_{aTDDB} . The TDDB equivalent-total-stress-time can be calculated as:

$$T_{TDDB}^E = \sum_{j=1}^J T_I \cdot \exp(\gamma_{TDDB} \cdot (V_g^j - V_D)) \cdot \exp\left(\frac{E_{aTDDB}}{K} \left(\frac{1}{T_S} - \frac{1}{T_j}\right)\right) \quad (6.11)$$

and the ESF is

$$ESF_{TDDB} = T_{TDDB}^E / T_A \quad (6.12)$$

2. HCD ESF

For HCD, lifetime dependence on voltage is modeled by the empirical expo-

ponential model with coefficient γ_{HCD} . The temperature dependence is modeled by the Arrhenius relationship with activation energy E_{aHCD} . The HCD equivalent-total-stress-time can be calculated as:

$$T_{HCD}^E = \sum_{j=1}^J T_I \cdot \exp(\gamma_{HCD} \cdot (\frac{1}{V_D} - \frac{1}{V_d^j})) \cdot \exp(\frac{E_{aHCD}}{K} (\frac{1}{T_S} - \frac{1}{T_j})) \quad (6.13)$$

and HCD ESF is:

$$ESF_{HCD} = T_{HCD}^E / T_A \quad (6.14)$$

3. NBTI ESF

NBTI degradation only happens to the PMOS transistor. Assume NBTI voltage dependence is modeled by the exponential model with voltage coefficient γ_{NBTI} , activation energy of temperature acceleration is E_{aNBTI} . NBTI equivalent-total-stress-time can be calculated as:

$$T_{NBTI}^E = \sum_{j=1}^J T_I \cdot \exp(\gamma_{NBTI} \cdot (V_g^j - V_D)) \cdot \exp(\frac{E_{aNBTI}}{K} (\frac{1}{T_S} - \frac{1}{T_j})) \quad (6.15)$$

and the ESF is:

$$ESF_{NBTI} = T_{NBTI}^E / T_A \quad (6.16)$$

For cell interconnect, EM ESF can be calculated by the same way. Assume interconnect W_0 has a stress current J_j in T_j , EM activation energy is E_{aEM} , and the standard stress current is J_S . From Black's equation and assume the current density power is n , the equivalent-total-stress-time can be calculated as:

$$T_{EM}^E = \sum_{j=1}^J T_I \cdot (J_j / J_S)^n \cdot \exp(\frac{E_{aEM}}{K} (\frac{1}{T_S} - \frac{1}{T_j})) \quad (6.17)$$

and the ESF is:

$$ESF_{EM} = T_{EM}^E / T_A \quad (6.18)$$

6.3.4.3 Cell Reliability

Cell transistor reliability can be estimated with the transistor operation profile, ESFs and failure mechanism lifetime distribution. For each failure mechanism, the best-fit lifetime distribution can be decided from acceleration tests.

Assume cell has an N type of operation, denote f_i as the frequency of the i th ($i = 1 \cdots N$) operation. F_0 is the device operating frequency. Transistor M_0 's ESFs of the i th operation are ESF_{TDDDB}^i , ESF_{HCD}^i and ESF_{NBTI}^i for TDDDB, HCD and NBTI, respectively. M_0 's reliability estimation takes following steps.

Individual Failure Mechanism

1. TDDDB

Weibull distribution is used to model TDDDB failures. Denote β as the Weibull distribution shape parameter, α_0 as the scale parameter under standard stress (V_D, T_S), M_0 's TDDDB reliability can be estimated by

$$R_{M_0}^{TDDDB}(t) = \exp\left[-\left(\frac{(\sum_{i=1}^N ESF_{TDDDB}^i \cdot f_i/F_0) \cdot t}{\alpha_0}\right)^\beta\right] \quad (6.19)$$

2. HCD

Lognormal distribution is normally used to model HCD failures. Denote μ_{HCD} and σ_{HCD} as the mean and sigma of the HCD lifetime distribution under standard stress, respectively. M_0 's HCD reliability can be estimated by

$$R_{M_0}^{HCD}(t) = 1 - \Phi\left(\frac{\ln((\sum_{i=1}^N ESF_{HCD}^i \cdot f_i/F_0) \cdot t) - \mu_{HCD}}{\sigma_{HCD}}\right) \quad (6.20)$$

3. NBTI

NBTI failures is often modeled by lognormal distribution. M_0 's NBTI reliability can be estimated by

$$R_{M_0}^{NBTI}(t) = 1 - \Phi\left(\frac{\ln((\sum_{i=1}^N ESF_{NBTI}^i \cdot f_i/F_0) \cdot t) - \mu_{NBTI}}{\sigma_{NBTI}}\right) \quad (6.21)$$

where μ_{NBTI} is the mean, σ_{NBTI} is the sigma of the lognormal distribution.

Transistor Reliability

If M_0 is an NMOS transistor, its reliability is

$$R_{M_0}(t) = R_{M_0}^{TDDDB}(t) \cdot R_{M_0}^{HCD}(t) \quad (6.22)$$

If M_0 is a PMOS transistor, and there is only NBTI degradation,

$$R_{M_0}(t) = R_{M_0}^{TDDDB}(t) \cdot R_{M_0}^{NBTI}(t) \quad (6.23)$$

Special attention must be paid for PMOS if NBTI and HCD coexist. Both NBTI and HCD cause threshold voltage degradation and the accumulation effect should be considered. With the ESFs, the mean of total threshold voltage degradation can be estimated by

$$\Delta V_{th}(t) = A_{HCD} \cdot \left(\sum_{i=1}^N ESF_{HCD}^i \cdot \frac{f_i}{F_0}\right) \cdot t^{n_{HCD}} + A_{NBTI} \cdot \left(\sum_{i=1}^N ESF_{NBTI}^i \cdot \frac{f_i}{F_0}\right) \cdot t^{n_{NBTI}} \quad (6.24)$$

where A_{HCD} and A_{NBTI} are prefactor of HCD and NBTI under the given standard stress conditions, respectively. n_{HCD} and n_{NBTI} is the power coefficient

of HCD and NBTI, respectively. The standard variation $\sigma_V(t)$ is

$$\sigma_V^2(t) = \sigma_{HCD}^2 \cdot \left(\sum_{i=1}^N ESF_{HCD}^i \cdot \frac{f_i}{F_0} \cdot t \right)^{2n_{HCD}} + \sigma_{NBTI}^2 \cdot \left(\sum_{i=1}^N ESF_{NBTI}^i \cdot \frac{f_i}{F_0} \cdot t \right)^{2n_{NBTI}} \quad (6.25)$$

where σ_{HCD} and σ_{NBTI} are the standard deviation of the prefactors of HCD and NBTI, respectively.

And PMOS reliability due to NBTI and HCD is

$$R_{M_0}^{NBTI+HCD}(t) = \Phi\left(\frac{V_{criterion} - \Delta V_{th}(t)}{\sigma_V(t)}\right) \quad (6.26)$$

where $V_{criterion}$ is the V_{th} degradation failure criterion.

Cell Interconnects Reliability

Reliability of cell interconnect W_0 can be estimated the same way. Lognormal distribution is applied with μ_{EM} as the mean and σ_{EM} as the sigma. ESF_{EM}^i is the ESF of the i th operation.

$$R_{W_0}^{EM}(t) = 1 - \Phi\left(\frac{\ln\left(\left(\sum_{i=1}^N ESF_{EM}^i \cdot f_i / F_0\right) \cdot t\right) - \mu_{EM}}{\sigma_{EM}}\right) \quad (6.27)$$

The cell reliability can be expressed as

$$R_{cell}(t) = \prod_{m=1}^M R_{M_m}(t) \cdot \prod_{w=1}^W R_{W_w}(t) \quad (6.28)$$

6.3.5 Chip Reliability Prediction

The chip's reliability can be expressed as

$$R_{Chip}(t) = R_{power}(t) \cdot \prod R_{block}(t) \quad (6.29)$$

where $R_{power}(t)$ is power network reliability, and $R_{block}(t)$ is functional block reliability.

6.3.5.1 Functional Block Reliability

For each functional block, reliability can be estimated by considering the block structure. In most cases, the functional block can be treated as a series system with cell as the component.

$$R_{block}(t) = \prod R_{cell}(t) \quad (6.30)$$

6.3.5.2 Power Network EM Estimation

To estimate power network EM, current waveform and interconnect temperature should be found at first. Current waveform can be obtained through SPICE simulation. Interconnect temperature can be estimated with Equ. 5.21 and Equ. 5.22.

For power network interconnect W_p ($p = 1 \cdots P$), stress is divided into L periods, current density in the l th ($l = 1 \cdots L$) stress period is J_p^l . Local interconnect temperature can be calculated by

$$T_p^l = T_S + \frac{J_p^{l2} \rho_0}{\frac{K_i}{\Theta t_i} [1 + 0.88 t_i / W] - J_p^{l2} \rho_0 \beta_m} \quad (6.31)$$

and the equivalent-total-stress-time under standard stress conditions is

$$T_{EM}^p = \sum_{l=1}^L T_l \cdot (J_p^l / J_S)^n \cdot \exp\left(\frac{E_{aEM}}{K} \left(\frac{1}{T_S} - \frac{1}{T_p^l}\right)\right) \quad (6.32)$$

W_p 's EM reliability can be estimated as

$$R_{W_p}^{EM}(t) = 1 - \Phi\left(\frac{\ln(T_{EM}^p) - \mu_{EM}}{\sigma_{EM}}\right) \quad (6.33)$$

Reliability of the power network is calculated by

$$R_{power}(t) = \prod_{p=1}^P R_{W_p}^{EM}(t) \quad (6.34)$$

6.4 Case Study: SRAM Reliability Prediction

SRAM is one of the most common circuit structures used in reliability simulation because it includes many typical subcircuits such as cross-connected six-transistor (6T) bit cell, precharge, decoding, and sense amplifier. Furthermore, SRAM is the most widely used on-chip memory [200]. For instance, a 1.1GHz 64-bit Sun UltraSPARC microprocessor has 87.5 million transistors, of which 63 million are in the SRAM cells [191]. The ever-increasing integration of SRAM in SoC design indicates that the reliability of modern VLSI systems depends on the reliability of on-chip memory. Lee et al. [1] did a large quantity of CPU and SRAM dynamic lifetests (about 3000 CPU and 6000 SRAM) and found 75% to 90% CPU failures occurred in the SRAM cache, although SRAM cache only occupies 50% of the CPU oxide area. They also found that near 90% of the SRAM failures were due to the memory array problem, the remaining 10% failure came from I/O circuits and the decoupling capacitor.

To demonstrate the PoF Statistical reliability prediction methodology, SRAM is selected as a vehicle in this case study. Without losing generality, reliability prediction is carried out for the memory core only.

6.4.1 SRAM Design

A 4096 words (16 bits/word) SRAM module is designed by using a commercial SRAM generator [201] that is based on IBM 90nm CMOS9SF process [202]. The SRAM operates at 1.2V, 500MHz. Gate oxide thickness of the bit cell is 1.4nm.

The power ring structure is used with a width of $25\mu\text{m}$.

In order to characterize the bit cell reliability, one SRAM bit cell with the peripheral control circuits are implemented. The cell has the same configuration of transistors as that from the generator. A cell schematic is shown in Fig.6.2. Transistors M1 - M4 form a latched structure for storing “1” or “0” at node “Store” depending on the differential voltages of BIT/BITn during write. The WORD line controls M5 and M6 and enables charging/discharging paths between the nodes Store/Storen and BIT/BITn lines during write/read cycles. The cell transfer ratio (width ratio of pass transistor to pull-down NMOS transistor, i.e., M5 to M1) is designed to be 1.5. The circuit block diagram used in SPICE simulation is shown

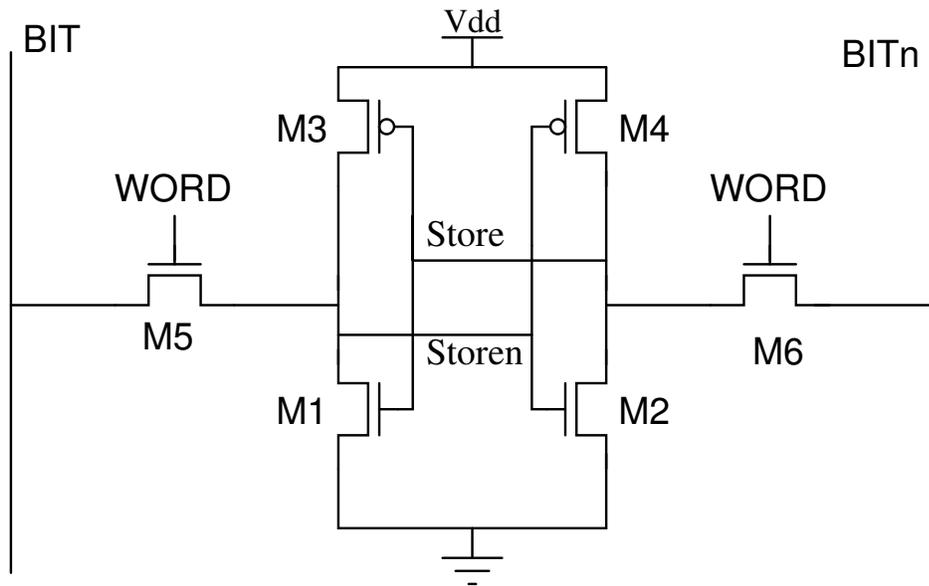


Figure 6.2: Schematic of one-bit 6T SRAM cell. Store/Storen represent cell state.

WORD line enables pass transistors M5 and M6 during memory read and rite.

in Fig. 6.3. Details of the peripheral circuits, including the precharge, read/write

control and the sense amplifier, can be found in [203].

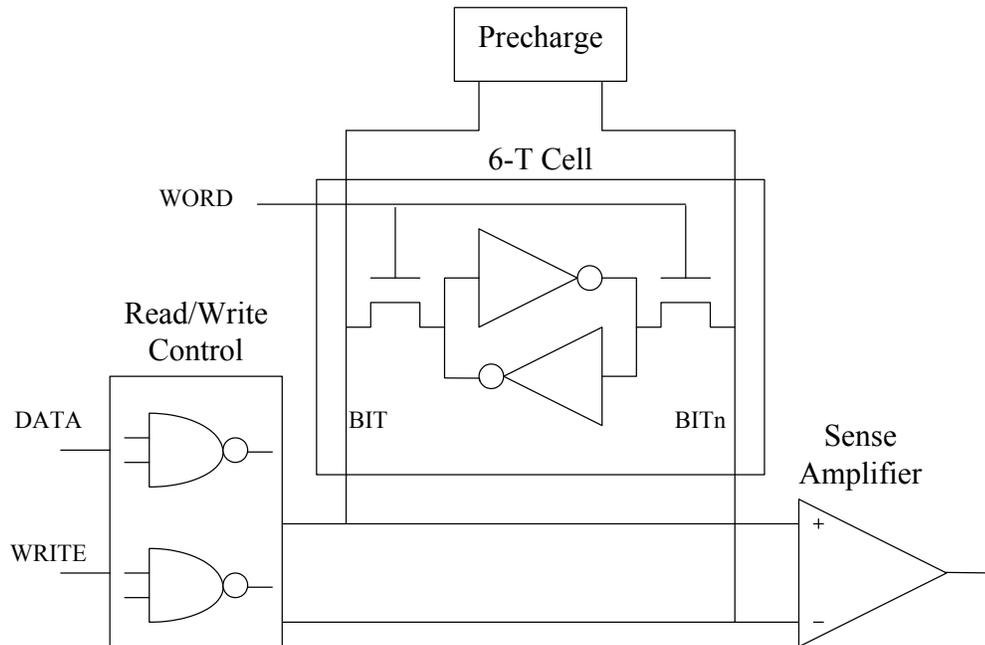


Figure 6.3: Block diagram of the simulation circuit in SRAM bit cell reliability characterization. The circuit consists of one 6T cell, read/write control logic and output sense amplifier.

There are three kinds of operation for SRAM cell, write, read and hold. The function of the cell is simulated in SPICE to perform a set of sequential “write 0, read 0, write 1, read 1” operations. Duration of each operation cycle is 2ns. The timing of input signals is given in Fig. 6.4.

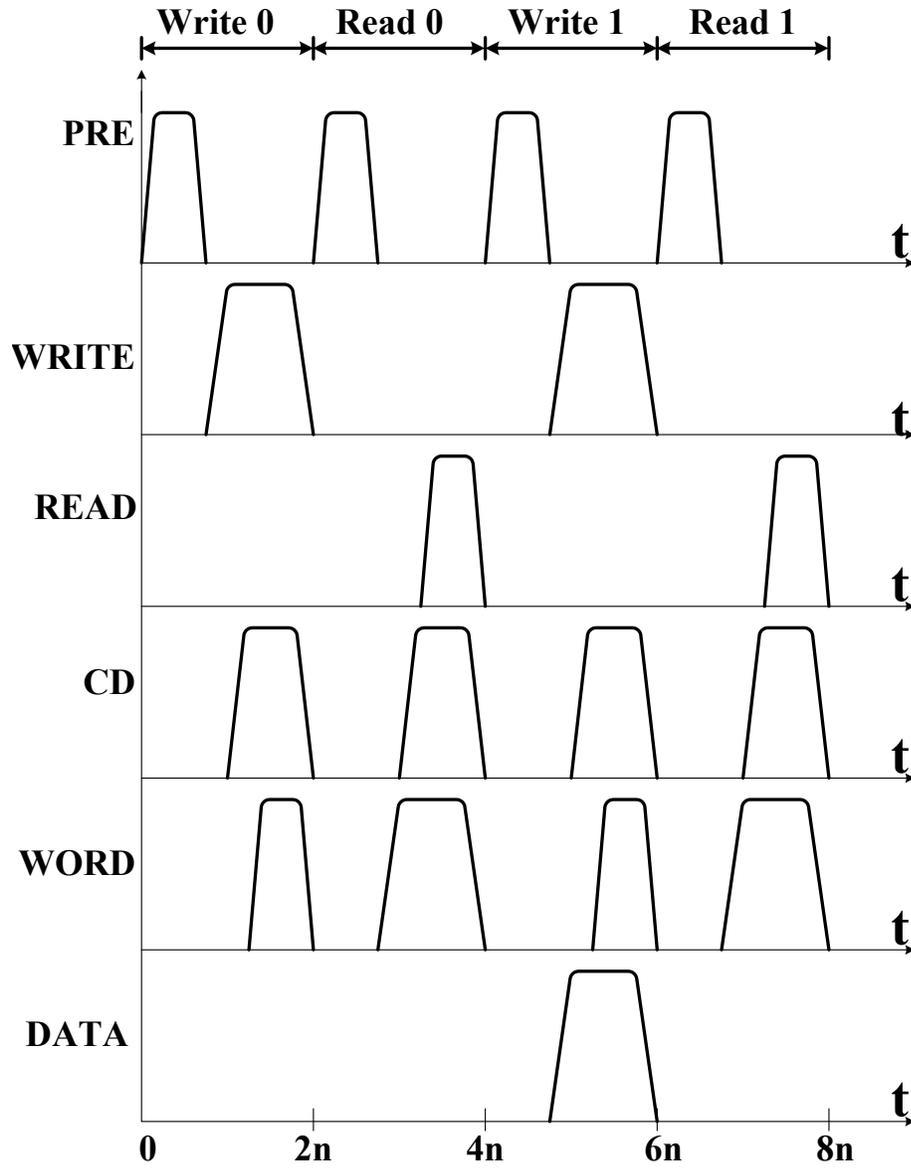


Figure 6.4: SRAM cell SPICE simulation stimuli. PRE is the precharge which exerts before each read/write operation. WORD signal controls cell transistor M5 and M6 during read and write. DATA carries the bit info “0” or “1” during each write operation.

6.4.2 Reliability Prediction

6.4.2.1 Application Profile

The ambient temperature $T_A = 25^\circ\text{C}$, operating voltage $V_D = 1.2\text{V}$, operation cycle $T_A = 2\text{ns}$. The SRAM module is running at maximum throughput at 500MHz with duty factor (DF) equals to one. All SRAM cells have the same opportunity of being read/write. For each cell, 50% operations are read and the rest 50% are write, data “0” and “1” are wrote to cell alternatively. All cells have the same operation profile as shown in Table 6.1.

Table 6.1: SRAM cell operation profile

Operation	write 0	read 0	write 1	read 1	hold 0	hold 1
Number/sec	30518	30518	30518	30518	2.50E+8	2.50E+8

6.4.2.2 Chip Thermal Analysis

Given the application profile, SRAM module’s power consumption can be estimated from SPICE simulation. In this case, the average power is 133.8mW. From Equ. 6.5, the substrate temperature T_S is found to be 33.5 °C.

6.4.2.3 SRAM Cell Reliability Characterization

The dynamic stress profile of the cell transistor is extracted from SPICE simulation output and post-processed by scripts (see Appendix A). Operation cycle $T_A = 2\text{ns}$, sampling interval T_I is set to 0.01ns. Failure mechanism voltage and

temperature lifetime models are listed in Table 6.2. Parameter values are not revealed for property issue.

Table 6.2: TDDB, NBTI and HCD lifetime models

	TDDB	NBTI	HCD
Voltage	$V_{gs}^{-\gamma_{TDDB}}$	$\exp(\frac{\gamma_{NBTI}}{V_{ds}})$	$\exp(-\gamma_{HCD} \cdot V_{gs})$
Temperature	$\exp(\frac{E_{aTDDB}}{kT})$	$\exp(\frac{E_{aNBTI}}{kT})$	$\exp(\frac{E_{aHCD}}{kT})$

ESFs are evaluated from the dynamic stress profile for all cell transistors.

Results are shown in Table 6.3.

Table 6.3: ESFs of SRAM cell transistors

	M1		M2		M3		M4		M5		M6	
	HCD	TDDB	HCD	TDDB	TDDB	NBTI	TDDB	NBTI	HCD	TDDB	HCD	TDDB
Write 0	0	0.6693	0	0.2741	0.0308	0.1106	0.6516	0.6619	0.0165	0.2801	0	0.2801
Read 0	0	0	0	1.0045	0.438	0.5975	0	0	0	0.5301	0	0.5301
Write 1	0	0.2742	0	0.6693	0.6516	0.6619	0.0308	0.1104	0	0.2801	0.0165	0.2801
Read 1	0	1.0043	0	0	0	0	0.4382	0.5976	0	0.5301	0	0.5301
Hold 0	0	0	0	1	1	1	0	0	0	0	0	0
Hold 1	0	1	0	0	0	0	1	1	0	0	0	0

To calculate cell transistors' total effective stress time of this application, another set of conversion factors, stress duty factors(SDFs), can be derived from ESFs and the cell operation profile. These SDFs are listed in Table 6.4. From the SDF results, it is obvious that TDDB is the most detrimental failure mechanism because M1, M2, M3 and M4 all have a SDF of 0.5 which means the gate oxide is under stress during half of the chip power-on time. NBTI must be dealt with carefully since M3 and M4 both suffer NBTI degradation with SDF=0.5. The two pass transistor, M5 and M6, only sustain very small TDDB and HCD degradation because of the small

Table 6.4: Cell transistor stress duty factors

	M1	M2	M3	M4	M5	M6
TDDDB	0.5	0.5	0.5	0.5	9.89E-5	9.89E-5
NBTI	0	0	0.5	0.5	0	0
HCD	0	0	0	0	1.01E-6	1.01E-6

number of read and write operations.

For each interconnect inside the cell, the current waveform was also obtained from SPICE simulation. The average current and the RMS current were calculated from the simulation output. Based on the simulation result, the ESFs for these cell interconnects are calculated to be 0.

6.4.2.4 Power Bus Reliability Estimation

The module's power bus has a ring structure and the ring width is 25 μm . Since the current density of the power bus is much larger than that of local cell interconnect, the local Joule heating effect must be considered in EM analysis. The average current of the power bus is 35.68mA. Using Equ.5.22 with parameters from [157], the temperature increase caused by Joule heating is estimated to be 19.1 $^{\circ}\text{C}$.

6.4.2.5 Memory Core Reliability Prediction

Weibull distribution is used to model TDDDB failures. For this SRAM module, the Weibull shape parameter is one that means gate oxide failure follows exponential distribution. Lognormal distribution is used to model HCD, NBTI and EM failures.

Lifetime model and distribution parameters can be found in [202, 54, 204]. For HCD, the failure criterion is $\Delta V_{th} \geq 25mV$. NBTI failure criterion is $|\Delta V_{th}| \geq 50mV$.

The SRAM core reliability prediction result is shown in Fig. 6.5. The reliability at the end of 20 years is estimated to be 0.9977. Failure rate estimation is shown in Fig. 6.6. It clearly shows that failure rate keeps at low value (below one FIT) till around 58,000 hours. After 58,000 hours, failure rate begins to increase. For the given application, HCD has a negligible effect because the read and write frequency is very low. Cell reliability simulation shows that TDDB and NBTI are the main concerns because they have larger ESF. Since this SRAM module only has 64K bit cells, oxide failure seldom happens because the small total gate oxide area. The estimated TDDB failure rate is 2.90E-4 FIT. NBTI also has a low failure rate at the beginning because of the degradation accumulation takes time. After 58,000 hours, NBTI begins to dominate and failure rate starts to increase. In this case, EM causes no failure because the power bus is wide and the memory size is small.

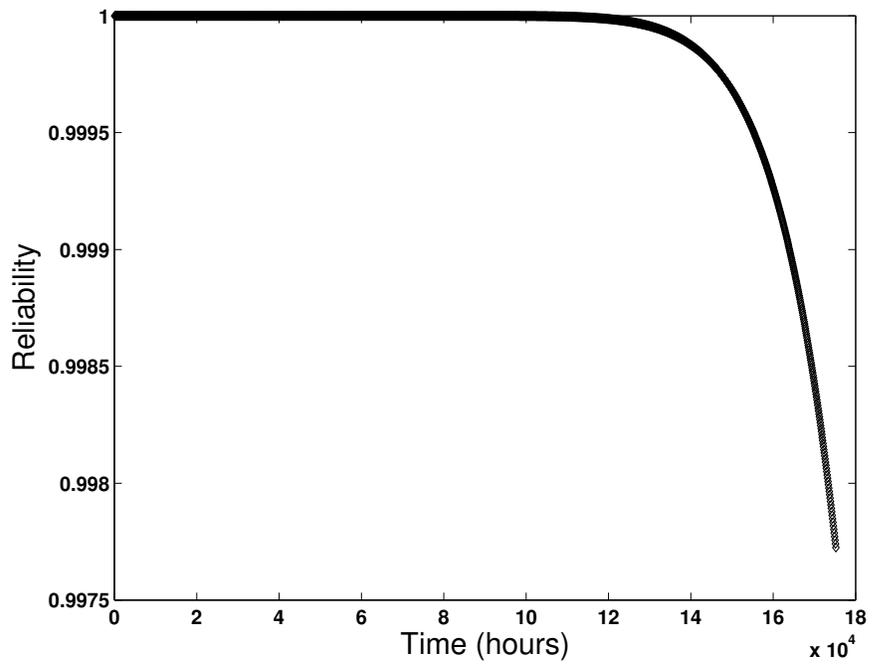


Figure 6.5: SRAM core reliability prediction.

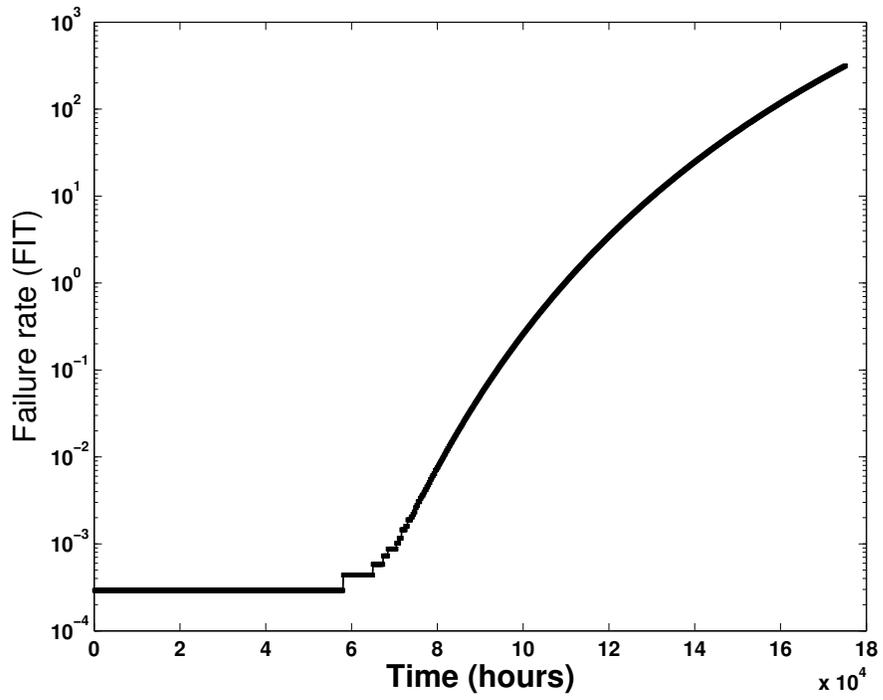


Figure 6.6: SRAM core failure rate prediction.

6.5 Summary

In this chapter, the conventional military handbook reliability prediction approach is briefly reviewed. A physics-of-failure based reliability prediction methodology is proposed to estimate device reliability in real application. The new methodology integrates physics-of-failure models into device reliability prediction together with lifetime distribution of each failure mechanism. The methodology is demonstrated by reliability prediction of a 90nm 64Kb SRAM core.

Chapter 7

Summary

7.1 Results

Driven by the consumer market which pursues performance and functionality, the semiconductor industry keeps introducing new materials and adapting new process technology to scale the device further down. The decreasing feature sizes, coupled with non-ideal voltage scaling, raises new reliability concerns such as NBTI and adversely affects those long-existing failure mechanisms: EM, HCD and TDDB. The increasing device complexity, narrowing the gap between normal operating and accelerated test conditions, makes it almost impossible or prohibitively expensive to qualify devices through accelerated tests with sufficient fault coverage. To take the leading role in technology and market without sacrificing reliability, manufacturers must be equipped with accurate reliability modeling and simulation tools to help control the reliability from the very beginning design stage to the final qualification.

Evidence in both recently published literature, and analyses of more than twenty years avionics failure data, signals imminent reliability challenge with increased failure rates and decreased wearout start time. Review of physics-of-failure and lifetime distributions of these intrinsic failure mechanisms reveal the scaling impact on device reliability. The multiple failure mechanisms era is approaching, as device is more vulnerable to defects because of the shrinking feature sizes. The con-

ventional military handbook based reliability prediction method, which is built on empirical data and incorporates all kinds of pre-factors, can't handle the challenge to predict device reliability more accurately in a reasonable time.

Lifetime distributions for failure mechanisms are important to reliability modeling and prediction because of the process variation. For EM, lognormal distribution have been widely used. Weibull distribution has been accepted to model TDDB failures. For HCD and NBTI, little research has been done to characterize their failure distributions. Lognormal distribution is often utilized without solid theoretical proof.

Based on the lifetime models and failure distributions, a new physics-of-failure based statistical reliability prediction methodology is proposed to handle the modeling and prediction challenges. The new methodology takes a unique top-down, bottom-up approach to reduce the modeling and simulation complexity. At the top level, two breakdowns are carried out in order to simplify the simulation. First is the IC structure breakdown in which VLSI circuit is divided into functional blocks and further the standard cells. Second is the application breakdown. A cell's operation profile is obtained through application analysis with regard to the device structure.

At the bottom level, a cell's reliability is characterized for all kinds of operation. With the SPICE simulator, the operation-based dynamic stress profile of cell transistors and interconnects are simulated and extracted by setting an appropriate sampling interval. The dynamic stress is further converted to an equivalent stress under a standard stress condition by utilizing the physics-of-failure based acceleration models. Those equivalent stresses are accumulated and used to estimate the

reliability of each mechanism based on the best-fit lifetime distribution. Cell reliability is estimated by competing failure mechanism model. The application-specific device reliability can be further predicted by considering the system structure.

A 90nm 64Kb SRAM module is designed and used as an example to demonstrate the prediction methodology. Dynamic voltage stresses of the SRAM bit cell during read, write and hold operations are extracted from SPICE simulation. For each failure mechanism, operation related ESFs are evaluated and used to discover the effective stress time. With the given application, simulation results showed that TDDB is the most serious reliability concern for the SRAM bit cell, NBTI is in the second place, and HCD has a negligible degradation effect. The memory core's reliability is also predicted to have a low constant failure rate before 58,000 hours, and increasing failure rate after that because NBTI wearout starts to kick in.

7.2 Future Work

Technology advancement always brings new reliability challenges. The work on accurate reliability modeling and prediction remains ongoing and this work is just a start. Plenty more work needs to be done on the road to achieve great reliability performance.

7.2.1 Model Verification and Validation

All the models used in the physics-of-failure based statistical approach are critical to accurate reliability modeling and prediction. Experimental work has to

be accomplished to verify and validate these models.

1. Physics-of-failure models.

Although many important research has been done to reveal the physics-of-failure process of failure mechanism, challenges still remain considering new material and processes are always being introduced in technology development. These changes may affect the degradation process and introduce new failure mechanism. The physics-of-failure models need to be updated following technology advancement.

2. Lifetime distribution.

The importance of lifetime distributions in reliability prediction can't be emphasized more strongly. Acceleration tests at circuit/device should be carried out with a large sample size.

7.2.2 Prediction Tool Development

System engineers are in great need of accurate reliability prediction tools. The physics-of-failure based statistical reliability prediction methodology provides a framework of prediction tools that will meet a system engineers' requirements by taking both the application and circuit into consideration. Based on the methodology, a product-specific reliability prediction tool can also be developed for device manufacturers.

Appendix A

Sample Scripts and Programs

A.1 Dynamic Stress Profile Extraction Script

This script is used to extract SRAM cell transistors' voltage and current from Cadence Spectre simulation output.

```
% Script

ocnWaveformTool( 'wavescan' )

simulator( 'spectre' )

design("/homes/qjin/cadence/simulation/SRAM/spectre/schematic/netlist/netlist")

resultsDir( "/homes/qjin/cadence/simulation/SRAM/spectre/schematic" )

modelFile('/users/qjin/home/cadence/SRAM/ibm013um.txt' '')

analysis('tran ?stop "8n" )

save( 'v "/WORD" "/vdd!" "/Storen" "/Store" "/BIT" "/BITn" )

save( 'i "/M1/B" "/M2/B" "/M6/B" "/M5/B" )

temp( 125 )

run()

selectResult( 'tran' )

plot(getData("/WORD") getData("/vdd!") getData("/Storen")
getData("/Store") getData("/BIT") getData("/M1/B") getData("/M2/B")
getData("/M6/B") getData("/M5/B") )
```

```

outputs()

out=outfile("./121v.out" "w")

for(tt 1 801

time=tt*0.00000000001-1*0.00000000001

fprintf(out "%4d " tt)

fprintf(out "%4.3f %4.3f %4.3f %4.3f %4.3f "

value(VT("/Storen"),time),value(VT("/Store"),time),value(VT("/WORD"),time),

value(VT("/BIT"),time),value(VT("/BITn"),time))

fprintf(out "%4.3f \n" value(VT("/vdd!"),time))

) close(out)

out1=outfile("./121C.out" "w")

for(tt 1 801

time=tt*0.00000000001-1*0.00000000001

fprintf(out1 "%4d " tt)

fprintf(out1 "%4.0f " value(IT("/M1/B"), time)*1000000000)

fprintf(out1 "%4.0f " value(IT("/M2/B"), time)*1000000000)

fprintf(out1 "%4.0f " value(IT("/M5/B"), time)*1000000000)

fprintf(out1 "%4.0f \n" value(IT("/M6/B"),time)*1000000000)

)

close(out1)

```

A.2 Example: MATLAB Script for Device Reliability Estimation

This MATLAB script is used to estimate SRAM cell reliability.

```
% Parameters need to be changed for different technology:

% AF parameters Vtn and T1

Vtn=0.0703491; %NMOS threshold voltage

M=801; %input rows

N=7; %input columns

T0=348.15; % normal temperature degree K

T1=273.15+25; %working temperature

% Voltage acceleration parameters

Btddb=12; Bhci=16;Bnbt=6; Bem=2;

% Activation energies

Etddb=0.9;Ehci=-0.2; Enbt=0.4; Eem=1.2;

%——TDDDB, HCI, NBTI———

filename='121v.out';

fid=fopen(filename,'r');

% Readin the voltages.

vtemp= fscanf(fid,'%d %g %g %g %g %g %g',[7 inf])

vtemp=vtemp';

fclose(fid);

for m = 1:M

for n = 1:N
```

```

if vtemp(m,n)<=0 vtemp(m,n) = 0;

end

end

end

% Calculate Vg,Vd for M1-M6

M1(:,1)=vtemp(:,3); M1(:,2)=vtemp(:,3); M1(:,3)=vtemp(:,2);

M2(:,1)=vtemp(:,2); M2(:,2)=vtemp(:,2); M2(:,3)=vtemp(:,3);

M3(:,1)=vtemp(:,3)-vtemp(:,7); M3(:,2)=vtemp(:,3)-vtemp(:,7);

M3(:,3)=vtemp(:,2)-vtemp(:,7); M4(:,1)=vtemp(:,2)-vtemp(:,7);

M4(:,2)=vtemp(:,2)-vtemp(:,7); M4(:,3)=vtemp(:,3)-vtemp(:,7);

M5(:,1)=vtemp(:,4);

for m=1:M

if vtemp(m,2)>=vtemp(m,5)

M5(m,2)=vtemp(m,4)-vtemp(m,5); M5(m,3)=vtemp(m,2)-vtemp(m,5);

else

M5(m,2)=vtemp(m,4)-vtemp(m,2); M5(m,3)=vtemp(m,5)-vtemp(m,2);

end

end

M6(:,1)=vtemp(:,4);

for m=1:M

if vtemp(m,3)>=vtemp(m,6)

M6(m,2)=vtemp(m,4)-vtemp(m,6);M6(m,3)=vtemp(m,3)-vtemp(m,6);

else

```

```

M6(m,2)=vtemp(m,4)-vtemp(m,3);M6(m,3)=vtemp(m,6)-vtemp(m,3);

end

end

% Calculate accumulated degradation

%NMOS M1

DM1=[0 0 0];

for m=1:M

if M1(m,1)>0 DM1(1)=DM1(1)+exp(Btddb*M1(m,1))*1;

end

if (M1(m,2)>Vtn)&(M1(m,3)>0) DM1(2)=DM1(2)+exp(-Bhci/M1(m,3))*1;

end

end

%NMOS M2

DM2=[0 0 0];

for m=1:M

if M2(m,1)>0 DM2(1)=DM2(1)+exp(Btddb*M2(m,1))*1;

end

if (M2(m,2)>Vtn)&(M2(m,3)>0) DM2(2)=DM2(2)+exp(-Bhci/M2(m,3))*1;

end

end

%NMOS M5

DM5=[0 0 0];

for m=1:M

```

```

if M5(m,1)>0 DM5(1)=DM5(1)+exp(Btddb*M5(m,1))*1;
end

if (M5(m,2)>Vtn)&(M5(m,3)>0) DM5(2)=DM5(2)+exp(-Bhci/M5(m,3))*1;
end

end

%NMOS M6

DM6=[0 0 0];

for m=1:M

if M6(m,1)>0 DM6(1)=DM6(1)+exp(Btddb*M6(m,1))*1;
end

if (M6(m,2)>Vtn)&(M6(m,3)>0) DM6(2)=DM6(2)+exp(-Bhci/M6(m,3))*1;
end

end

%PMOS M3 Negative gate voltage

DM3=[0 0 0];

for m=1:M

if M3(m,1)<0

DM3(1)=DM3(1)+exp(-Btddb*M3(m,1))*1;DM3(3)=DM3(3)+exp(Bnbt*(-M3(m,1)))*1;
end

end

%PMOS M4

DM4=[0 0 0];

for m=1:M

```

```

if M4(m,1)<0
DM4(1)=DM4(1)+exp(-Btddb*M4(m,1))*1;DM4(3)=DM4(3)+exp(Bnbti*(-M4(m,1)))*1;
end
end
% Degradation index
DT=[0 0 0 0]; temp1=[DM1(2) DM2(2) DM3(2) DM4(2) DM5(2) DM6(2)];
temp2=[DM1(3) DM2(3) DM3(3) DM4(3) DM5(3) DM6(3)];
DT(1)=(DM1(1)+DM2(1)+DM3(1)+DM4(1)+DM5(1)+DM6(1))*exp(-Etddb*11605*(1/T1));
DT(2)=max(temp1)*exp(-Ehci*11605*(1/T1));
DT(3)=max(temp2)*exp(-Enbti*11605*(1/T1));
%---EM-----
filename='121C.out';
fid=fopen(filename,'r');
% Readin the voltages.
% vtemp(:,1):time stamp,vtemp(:,2):Storen,
% vtemp(:,3):Store,vtemp(:,4):WORD,vtemp(:,5):BIT,
% vtempitemp= fscanf(fid,'itemp=itemp');
fclose(fid);
% Vdd GND current
Idd=itemp(:,2)+itemp(:,3);
Iss=itemp(:,4)+itemp(:,5);
DEM=[0 0];
for m=1:M

```

```
if Idd(m)>0 DEM(1) = DEM(1) + Idd(m)2; end
if Iss(m)<0 DEM(2) = DEM(2) + Iss(m)2; end
end
DT(4)=(DEM(1)+DEM(2))*exp(-Eem*11605/T1);
```

Bibliography

- [1] Y. H. Lee, N. Mielke, M. Agostinelli, S. Gupta, R. Lu, and W. McMahon, "Prediction of logic product failure due to thin-gate oxide breakdown," in *IRPS*, pp. 18–28, 2006.
- [2] T. J. Anderson and J. M. Carulli Jr., "Modeling and monitoring of product DPPM with multiple fail modes," in *IRPS*, pp. 545–551, 2006.
- [3] J. Qin, B. Huang, and J. Bernstein, "SDRAM test report." AVSI #17, 2005.
- [4] M. Ohring, *Reliability and Failure of Electronic Materials and Devices*. Academic Press, 1998.
- [5] R. Degraeve, B. Kaczer, and G. Groesenken, "Degradation and breakdown in thin oxide layers: mechanisms, models and reliability prediction," *Microelectronics Reliability*, vol. 39, pp. 1445–1460, 1999.
- [6] J. H. Stathis, "Reliability limits for the gate insulator in CMOS technology," *IBM Journal of Research and Development*, vol. 46, no. 2/3, pp. 265–286, 2002.
- [7] E. Y. Wu and J. Suñé, "Power-law voltage acceleration: a key element for ultra-thin gate oxide reliability," *Microelectronics Reliability*, vol. 45, pp. 1809–1834, 2005.
- [8] K. O. Jeppson and C. M. Svensson, "Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices," *Journal of Applied Physics*, vol. 48, pp. 2004–2014, 1977.
- [9] A. Acovic, G. L. Rosa, and Y.-C. Sun, "A review of hot carrier degradation mechanisms in MOSFETs," *Microelectronics Reliability*, vol. 36, pp. 845–869, 1996.
- [10] D. Young and A. Christou, "Failure mechanism models for electromigration," *IEEE Transactions on Reliability*, vol. 43, pp. 186–192, June 1994.
- [11] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, pp. 164–173. Cambridge University Press, 1998.
- [12] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers, "The case for lifetime reliability-aware microprocessor," in *IEEE Proceedings of 31st Annual International Symposium on Computer Architecture*, pp. 276–287, 2004.
- [13] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 1998.

- [14] D. J. Dumin, ed., *Oxide Reliability: A Summary of Silicon Oxide Wearout, Breakdown, and Reliability*. World Scientific Publishing, 2002.
- [15] J. H. Stathis and D. J. DiMaria, “Reliability projection for ultra-thin oxides at low voltage,” in *IEDM*, pp. 167–170, 1998.
- [16] D. J. DiMaria and J. H. Stathis, “Explanation for the oxide thickness dependence of breakdown characteristics of metal-oxide-semiconductor structures,” *Applied Physics Letters*, vol. 70, pp. 2708–2710, 1997.
- [17] J. H. Stathis, “Percolation models for gate oxide breakdown,” *Journal of Applied Physics*, vol. 86, pp. 5757–5766, 1999.
- [18] E. Y. Wu, J. Sune, and W. Lai, “On the weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination—part II: Experimental results and the effects of stress conditions,” *IEEE Transactions on Electron Devices*, vol. 49, pp. 2141–2150, 2002.
- [19] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, “Characteristics of the surface-state charge of thermally oxidized silicon,” *Journal of the Electrochemical Society*, vol. 114, pp. 266–274, 1967.
- [20] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, “A through investigation of MOSFETs NBTI degradation,” *Microelectronics Reliability*, vol. 45, pp. 83–89, 2005.
- [21] D. K. Schroder and J. A. Babcock, “Negative bias temperature instability: Road to cross in deep submicron semiconductor manufacturing,” *Journal of Applied Physics*, vol. 94, pp. 1–18, 2003.
- [22] R. Blish, T. Dellin, S. Huber, M. Johnson, J. Maiz, B. Likins, N. Lycoudes, J. PcPherson, Y. Peng, C. Peridier, A. Preussger, G. Prokop, and L. Tullos, “Critical reliability challenges for the international technology roadmap for semiconductor (ITRS),” in *Technology Transfer 03024377A-TR*, International SEMATECH, 2003.
- [23] J. Qin and J. Bernstein, “Non-Arrhenius temperature acceleration and stress-dependent voltage acceleration for semiconductor device involving multiple failure mechanisms,” in *IIRW*, pp. 93–97, 2006.
- [24] *MIL-HDBK-217F Reliability prediction of electronic equipment*. Washington DC 20301: Department of Defense, December 1991.
- [25] J. A. McInn, “Constant failure rate—a paradigm in transition?,” *Quality and Reliability Engineering International*, vol. 6, pp. 237–241, 1990.
- [26] R. F. Drenick, “The failure law of complex equipment,” *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, pp. 680–690, Dec 1960.

- [27] J. B. Bowles, "A survey of reliability-prediction procedures for microelectronic devices," *IEEE Transactions on Reliability*, vol. 41, pp. 2–12, March 1992.
- [28] Cadence Design Systems, Inc, *Virtuoso UltraSim Simulator User Guide*, 6.0 ed., Nov 2005.
- [29] M. Karam, W. Fikry, H. Haddara, and H. Ragai, "Implementation of hot carrier reliability simulation in Eldo," in *IEEE International Symposium on Circuits and Systems*, pp. 515–518, 2001.
- [30] J. Qin, B. Huang, M. Talmor, X. Li, X. Zhang, Z. Gur, and J. Bernstein, "Lifetime analysis of avionics in commercial aerospace industry," in *Components for Military and Space Electronics*, pp. 196–200, 2005.
- [31] H. Iwai, H. S. Momose, M. Saito, M. Ono, and Y. Katsumata, "The future of ultra-small-geometry MOSFETs beyond 0.1 micron," *Microelectronic Engineering*, vol. 28, pp. 147–154, 1995.
- [32] J. S. Suehle, "Ultrathin gate oxide reliability: physical models, statistics, and characterization," *IEEE Transactions on Electron Devices*, vol. 49, pp. 958–971, 2002.
- [33] R. Degraeve, B. Kaczer, M. Houssa, G. Groeseneken, and M. Heyns, "Analysis of high voltage TDDB measurements on $t_{a2}O_5SiO_2$ stack," in *IEDM*, pp. 13.4.1–13.4.3, 1999.
- [34] S.-H. Lo, D. A. Buchanan, and Y. Taur, "Modeling and characterization of quantization, polysilicon depletion and direct tunnelling effects in MOSFETs with ultrathin oxides," *IBM Journal of Research and Development*, vol. 43, pp. 327–337, 1999.
- [35] J. C. Ranuáñez, M. J. Deen, and C.-H. Chen, "A review of gate tunneling current in MOS devices," *Microelectronics Reliability*, vol. 46, pp. 1939–1956, 2006.
- [36] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO₂ thin films," *Journal of Applied Physics*, vol. 84, pp. 1513–1523, 1998.
- [37] J. H. Stathis, "Physical and predictive models of ultrathin oxide reliability in CMOS devices and circuits," *IEEE Transactions on Device and Materials Reliability*, vol. 1, pp. 43–59, 2001.
- [38] E. M. Vogel, J. S. Suehle, M. D. Edelstein, B. Wang, Y. Chen, and J. B. Bernstein, "Reliability of ultrathin silicon dioxide under combined substrate hot-electron and constant voltage tunneling," *IEEE Transactions on Electron Devices*, vol. 47, pp. 1183–1191, 2000.

- [39] K. F. Schuegraf and C. Hu, "Hole injection SiO₂ breakdown model for very low voltage lifetime extrapolation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 761–767, 1994.
- [40] J. F. Verweij and J. H. Klootwijk, "Dielectric breakdown I: A review of oxide breakdown," *Microelectronics*, vol. 27, pp. 611–622, 1996.
- [41] JEDEC Publication JEP122B, *Failure Mechanisms and Models for Semiconductor Devices*, 2003.
- [42] M. V. Fischetti, "Model for the generation of positive charge at the Si-SiO₂ interface based on hot-hole injection from the anode," *Physical Review B*, vol. 31, pp. 2099–2113, 1985.
- [43] D. J. DiMaria, E. Cartier, and D. A. Buchanan, "Anode hole injection and trapping in silicon dioxide," *Journal of Applied Physics*, vol. 80, pp. 304–317, 1996.
- [44] M. Rasras, I. D. Wolf, G. Groeseneken, B. Kaczer, R. Degraeve, and H. E. Maes, "Photo-carrier generation as the origin of Fowler-Nordheim-induced substrate hole current in thin oxides," *IEEE Transactions on Electron Devices*, vol. 48, no. 2, pp. 231–238, 2001.
- [45] D. J. DiMaria and J. W. Stasiak, "Trap creation in silicon dioxide produced by hot electrons," *Journal of Applied Physics*, vol. 65, pp. 2342–2356, 1988.
- [46] D. J. DiMaria, "Electron energy dependence of metal-oxide-semiconductor degradation," *Applied Physics Letters*, vol. 75, no. 16, pp. 2427–2428, 1999.
- [47] K. Hess, I. C. Kizilyalli, and J. W. Lyding, "Giant isotope effect in hot electron degradation of metal oxidesilicon devices," *IEEE Transactions on Electron Devices*, vol. 45, no. 2, pp. 406–416, 1998.
- [48] J. Wu, E. Rosenbaum, B. MacDonald, E. Li, J. Tao, B. Tracy, and P. Fang, "Anode hole injection versus hydrogen release: The mechanism for gate oxide breakdown," in *IRPS*, (San Jose, CA USA), pp. 27–32, 2000.
- [49] J. S. Suehle, P. Chaparala, C. Messick, W. M. Miller, and K. C. Boyko, "Field and temperature acceleration of time-dependent dielectric breakdown in intrinsic thin SiO₂," in *IEEE IRPS*, pp. 120–125, 1994.
- [50] C. Hu and Q. Lu, "A unified gate oxide reliability model," in *IRPS*, pp. 47–51, 1999.
- [51] P. E. Nicollian, W. R. Hunter, and J. C. Hu, "Experimental evidence for voltage driven breakdown models in ultrathin gate oxides," in *IRPS*, pp. 7–15, 2000.

- [52] D. J. Bude, B. E. Weir, and P. J. Silverman, "Explanation of stress-induced damage in thin oxides," in *IEDM*, pp. 179–182, 1998.
- [53] J. M. McKenna, E. Y. Wu, and S.-H. Lo, "Tunneling current characteristics and oxide breakdown in P+ polygate PFET capacitors," in *IRPS*, pp. 16–20, 2000.
- [54] E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, and J. McKenna, "Voltage-dependent voltage-acceleration of oxide breakdown for ultra-thin oxides," in *IEDM*, pp. 22.4.1–22.4.4, 2000.
- [55] K. Ohgata, M. Ogasawara, K. Shiga, S. Tsujikawa, E. Murakami, H. Kato, H. Umeda, and K. Kubota, "Universality of power-law voltage dependence for TDDB lifetime in thin gate oxide PMOSFETs," in *IRPS*, pp. 372–376, 2005.
- [56] R. Duschl and R.-P. Vollertsen, "Is the power-law model applicable beyond the direct tunneling regime?," *Microelectronics Reliability*, vol. 45, pp. 1861–1867, 2005.
- [57] T. Pompl and M. Röhner, "Voltage acceleration of time-dependent breakdown of ultra-thin gate dielectrics," *Microelectronics Reliability*, vol. 45, pp. 1835–1841, 2005.
- [58] D. J. Dumin, S. K. Mopuri, S. Vanchinathan, R. S. Scott, R. Subramoniam, and T. G. Lewis, "High field related thin oxide wearout and breakdown," *IEEE Transactions on Electron Devices*, vol. 42, pp. 760–772, 1995.
- [59] J. S. Suehle and P. Chaparala, "Low electric field breakdown of thin SiO_2 films under static and dynamic stress," *IEEE Transactions on Electron Devices*, vol. 44, no. 5, pp. 801–808, 1997.
- [60] M. Kimura, "Oxide breakdown mechanism and quantum physical chemistry for time-dependent dielectric breakdown," in *35th IEEE IRPS*, (Denver, CO USA), pp. 190–200, 1997.
- [61] A. Yassine, H. E. Nariman, and K. Olasupo, "Field and temperature dependence of TDDB of ultrathin gate oxide," *IEEE Electron Device Letters*, vol. 20, pp. 390–392, 1999.
- [62] J. McPherson, V. Reddy, K. Banerjee, and H. Le, "Comparison of E and 1/E TDDB models for SiO_2 under long-term/low-field test conditions," in *IEDM*, pp. 7.3.1–7.3.4, 1998.
- [63] E. Y. Wu, A. Vayshenker, E. Nowak, J. Sune, R. P. Vollertsen, W. Lai, and D. Harmon, "Experimental evidence of t_{bd} power-law for voltage dependence of oxide breakdown in ultrathin gate oxides," *IEEE Transactions on Electron Devices*, vol. 49, pp. 2244–2253, 2002.

- [64] R. Duschl and R.-P. Vollertsen, "Voltage acceleration of oxide breakdown in the sub-10nm Fowler-Nordheim and direct tunneling regime," in *IRW*, pp. 44–48, IEEE, 2005.
- [65] J. W. McPherson and H. C. Mogul, "Distributed bonding states in SiO₂ thin-films and their impact on TDDB," in *IRPS*, pp. 47–56, 1998.
- [66] J. S. Suehle, E. M. Vogel, B. Wang, and J. B. Bernstein, "Temperature dependence of soft breakdown and wearout in sub 3 nm SiO₂ films," in *IRPS*, pp. 33–39, 2000.
- [67] R. Moazzami, J. C. Lee, and C. Hu, "Temperature acceleration of time-dependent dielectric breakdown," *IEEE Transactions on Electron Devices*, vol. 36, pp. 2462–2465, 1989.
- [68] D. J. DiMaria and J. H. Stathis, "Non-Arrhenius temperature dependence of reliability in ultrathin silicon dioxide films," *Applied Physics Letters*, vol. 74, pp. 1752–1754, 1999.
- [69] B. Kaczer, R. Degraeve, N. Pagon, T. Nigam, and G. Groesenken, "Investigation of temperature acceleration of thin oxide time-to-breakdown," *Microelectronic Engineering*, vol. 48, no. 1-4, pp. 47–50, 1999.
- [70] R. Degraeve, N. Pagon, B. Kaczer, T. Nigam, G. Groeseneken, and A. Naem, "Temperature acceleration of oxide breakdown and its impact on ultra-thin gate oxide reliability," in *1999 Symposium on VLSI Technology Digest of Technical Papers*, pp. 59–60, 1999.
- [71] 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.
- [72] E. Wu, J. Sune, and W. Lai, "Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin oxides," *Microelectronic Engineering*, vol. 59, pp. 25–31, 2001.
- [73] 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, pp. 904–911, 1998.
- [74] J. Sune, I. Placencia, N. Barniol, E. Farres, F. Martin, and X. Aymerich, "On the breakdown statistics of very thin S_iO_2 films," *Thin Solid Films*, vol. 185, pp. 347–362, 1990.
- [75] E. Y. Wu and R. P. Vollertsen, "On the weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination—part I: Theory, methodology, experimental techniques," *IEEE Transactions on Electron Devices*, vol. 49, pp. 2131–2140, 2002.

- [76] J. Suñé, “New physics-based analytic approach to the thin-oxide breakdown statistics,” *IEEE Electron Device Letters*, vol. 22, pp. 296–298, 2001.
- [77] E. Y. Wu, W. W. Abadeer, L.-K. Han, S.-H. Lo, and G. R. Hueckel, “Challenges for accurate reliability projections in the ultrathin oxide regime,” in *IRPS*, pp. 57–65, 1999.
- [78] B. E. Deal, “The current understanding of charges in the thermally oxidized silicon structure,” *Journal of the Electrochemical Society*, vol. 121, pp. 198c–205c, 1974.
- [79] A. Goetzberger, A. D. Lopez, and R. J. Strain, “On the formaton of surface states during stress aging of thermal Si-SiO₂ interfaces,” *Journal of the Electrochemical Society*, vol. 120, pp. 90–96, 1973.
- [80] M. A. Alam and S. Mahapatra, “A comprehensive model of PMOS NBTI degradation,” *Microelectronics Reliability*, vol. 45, pp. 71–81, 2005.
- [81] V. Huard, M. Denais, and C. Parthasarathy, “NBTI degradation: From physical mechanisms to modeling,” *Microelectronics Reliability*, vol. 46, pp. 1–23, 2006.
- [82] J. H. Stathis and S. Zafar, “The negative bias temperature instability in MOS devices: a review,” *Microelectronics Reliability*, vol. 46, pp. 270–286, 2006.
- [83] D. K. Schroder, “Negative bias temperature instability: what we do understand?,” *Microelectronics Reliability*, vol. 47, pp. 841–852, 2007.
- [84] M. A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, “A comprehensive model for PMOS NBTI degradation: recent progress,” *Microelectronics Reliability*, vol. 47, pp. 853–862, 2007.
- [85] S. Ogawa and N. Shiono, “Generalized diffusion-reaction model for the low-field charge-buildup instability at the si-SiO₂ interface,” *Physical Review B*, vol. 51, no. 7, pp. 4218–4230, 1995.
- [86] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, “Proton-induced defect generation at the Si-SiO₂ interface,” *IEEE Transactions on Nuclear Science*, vol. 48, pp. 2086–2092, 2001.
- [87] W. Abadeer and W. Ellis, “Behavior of NBTI under AC dynamic circuit conditions,” in *IRPS*, pp. 17–22, 2003.
- [88] S. Rangan, N. Mielke, and E. C. Yeh, “Universal recovery behavior of negative bias temperature instability,” in *IEDM*, pp. 341–344, IEEE, 2003.
- [89] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. Kwong, “Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling,” *IEEE Electron Device Letters*, vol. 23, pp. 734–736, 2002.

- [90] D. Varghese, S. Mahapatra, and M. A. Alam, "Hole energy dependent interface trap generation in MOSFET Si/SiO₂ interface," *IEEE Electron Device Letters*, vol. 26, pp. 572–574, 2005.
- [91] B. Zhu, J. S. Suehle, and J. B. Bernstein, "Mechanism for reduced NBTI effect under pulsed bias stress conditions," in *IRPS*, pp. 689–690, 2004.
- [92] S. Zafar, B. H. Lee, and J. Stathis, "Evaluation of NBTI in HfO₂ gate-dielectric stacks with tungsten gates," *IEEE Electron Device Letters*, vol. 25, no. 3, pp. 153–155, 2004.
- [93] S. Chakravarthi, A. T. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature instability," in *42nd IRPS*, pp. 273–282, IEEE, 2004.
- [94] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI degradation and its impact on electric field dependence of the lifetime," in *IRPS*, pp. 23–27, IEEE, 2004.
- [95] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation," *Journal of Applied Physics*, vol. 97, pp. 1–9, 2005.
- [96] G. Haller, M. Knoll, D. Braunig, F. Wulf, and W. R. Fahrner, "Bias-temperature stress on metal-oxide-semiconductor structures as compared to ionizing irradiation and tunnel injection," *Journal of Applied Physics*, vol. 56, p. 184, 1984.
- [97] P. Chaparala, J. Shibley, and P. Lim, "Threshold voltage drift in PMOSFETS due to NBTI and HCI," in *IRW*, pp. 95–97, IEEE, 2000.
- [98] S. Mahapatra, P. B. Kumar, and M. A. Alam, "Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs," *IEEE Transactions on Electron Devices*, vol. 51, no. 9, pp. 1371–1379, 2004.
- [99] E. Takeda, C. Y. Yang, and A. Miura-Hamada, *Hot-Carrier Effects in MOS Devices*. Academic Press, 1995.
- [100] J. J. Sanchez and T. A. DeMassa, "Review of carrier injection in the silicon/silicon-dioxide system," in *IEE Proceedings*, vol. 138, pp. 377–389, 1991.
- [101] P. E. Cottrell, R. R. Troutman, and T. H. Ning, "Hot-electron emission in n-channel IGFETs," *IEEE Journal of Solid-State Circuits*, vol. 14, no. 2, pp. 442–455, 1979.
- [102] T. H. Ning, "Hot -electron emission currents in n-channel IGFET's," in *International Electron Devices Meeting*, pp. 144–147, IEEE, 1977.

- [103] N. Koike and K. Tatsuuma, "A drain avalanche hot carrier lifetime model for n- and p-channel MOSFETs," *IEEE Transactions on Device and Materials Reliability*, vol. 4, no. 3, pp. 457–466, 2004.
- [104] D. A. Buchanan, "On the generation of interface states from electron-hole recombination in metal-oxide-semiconductor capacitors," *Applied Physics Letters*, vol. 65, pp. 1257–1259, 1994.
- [105] W. Shockley, "Problems related to p-n junctions in silicon," *Solid-State Electron*, vol. 2, pp. 35–67, 1961.
- [106] J. F. Verwey, R. P. Kramer, and B. J. de Maagt, "Mean free path of hot electrons at the surface of boron-doped silicon," *Journal of Applied Physics*, vol. 46, pp. 2612–2619, 1975.
- [107] S. Tam, P.-K. Ko, and C. Hu, "Lucky-electron model of channel hot-electron injection in MOSFET's," *IEEE Transactions on Electron Devices*, vol. ED-31, pp. 1116–1125, 1984.
- [108] C. Hu, "Lucky-electron model of channel hot electron emission," in *IEDM Technical Digest*, pp. 22–25, 1979.
- [109] C. Hu, S. C. Tam, F.-C. Hsu, P.-K. Ko, T.-Y. Chan, and K. W. Terrill, "Hot-electron-induced MOSFET degradation-model, monitor, and improvement," *IEEE Journal of Solid-State Circuits*, vol. SC-20, pp. 295–305, 1985.
- [110] K. Hasnat, C.-F. Yeap, S. Jallepalli, W.-K. Shih, S. A. Hareland, V. M. A. Jr, A. F. T. Jr, and C. M. Maziar, "A pseudo-lucky electron model for simulation of electron gate current in submicron NMOSFET's," in *IEEE Transactions on Electron Devices*, pp. 1264–1273, 1996.
- [111] P. K. Ko, R. S. Muller, and C. Hu, "A unified model for hot-electron currents in MOSFET's," in *International Electron Devices Meeting*, pp. 600–603, IEEE, 1981.
- [112] Y. Leblebici and S. M. Kang, *Hot Carrier Reliability of MOS VLSI Circuits*, p. 61. Kluwer Academic Publisher, 1993.
- [113] B. S. Doyle and K. R. Mistry, "A lifetime prediction method for hot carrier degradation in surface-channel p-MOS devices," *IEEE Transactions on Electron Devices*, vol. 37, pp. 1301–1307, 1990.
- [114] R. Woltjer, A. Hamada, and E. Takeda, "Time dependence of p-MOSFET hot carrier degradation measured and interpreted consistently over ten orders of magnitude," *IEEE Transactions on Electron Devices*, vol. 40, no. 2, pp. 392–401, 1993.

- [115] E. Sangiorgi, "Historical perspective and recent developments of hot-carrier generation modeling for device analysis," in *International Conference on Simulation of Semiconductor Processes and Devices*, pp. 5–8, 1997.
- [116] B. Meinershagen, "Consistent gate and substrate current modeling based on energy transport and the lucky electron concept," in *IEDM*, pp. 504–507, 1988.
- [117] T. Toyabe and H. Kodera, "A theory for inter-valley transfer effect in two-valley semiconductor," *Jpn. J. Appl. Phys.*, vol. 13, pp. 1404–1413, 1974.
- [118] E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection," *IEEE Electron Device Letters*, vol. EDL-4, pp. 111–113, 1983.
- [119] E. Takeda, R. Izawa, K. Umeda, and R. Nagai, "AC hot-carrier effects in scaled MOS devices," in *IRPS*, (Las Vegas, NV USA), pp. 118–122, 1991.
- [120] K. N. Quader, P. Fang, and J. T. Yue, "Hot carrier reliability design rules for translating device degradation to CMOS digital circuit degradation," *IEEE Transactions on Electron Devices*, vol. 41, pp. 681–691, 1994.
- [121] P. Heremans, G. V. D. Bosch, R. Bellens, G. Groseneken, and H. E. Maes, "Understanding of the temperature dependence of channel hot-carrier degradation in the range 77 k to 300 k," in *IEDM*, pp. 67–70, IEEE, 1989.
- [122] M. Song, K. P. MacWilliams, J. S. Cable, and J. C. Woo, "Bias and temperature dependence of hot carrier lifetime from 77k to 300k," in *IEDM*, pp. 707–780, 1992.
- [123] E. S. Snyder, A. Kapoor, and C. Anderson, "The impact of statistics on hot-carrier lifetime estimates of n-channel MOSFETs," in *SPIE - Microelectronics Manufacturing and Reliability*, vol. 1802, pp. 180–187, 1992.
- [124] H. Kim and H. Hwang, "Effect of statistical process variation of MOSFET on hot carrier lifetime," *Solid-State Electronics*, vol. 43, pp. 989–991, 1999.
- [125] A. Kuntman, A. Ardah, H. Kuntman, and F. Kacar, "A Weibull distribution-based new approach to represent hot carrier degradation in threshold voltage of MOS transistors," *Solid-State Electronics*, vol. 48, pp. 217–223, 2004.
- [126] D. G. Pierce and P. G. Brusius, "Electromigration: A review," *Microelectron Reliability*, vol. 37, pp. 1053–1072, 1997.
- [127] J. R. Black, "Electromigration failure modes in aluminum metallization for semiconductor devices," *Proceedings of the IEEE*, vol. 57, no. 9, pp. 1587–1594, 1969.
- [128] A. Scorzoni, B. Neri, C. Caprile, and F. Fantini, "Electromigration in thin-film interconnection lines: models, methods and results," *Materials Science Reports*, vol. 7, pp. 143–220, 1991.

- [129] D. W. Malone and R. E. Hummel, "Electromigration in integrated circuits," *Critical Reviews in Solid State and Materials Science*, vol. 22, no. 3, pp. 199–238, 1997.
- [130] E. T. Ogawa, K.-D. Lee, V. A. Blaschke, and P. S. Ho, "Electromigration reliability issues in dual-damascene Cu interconnections," *IEEE Transactions on Reliability*, vol. 51, no. 4, pp. 403–419, 2002.
- [131] K. N. Tu, "Recent advances on electromigration in very-large-scale-integration of interconnects," *Journal of Applied Physics*, vol. 94, pp. 5451–5473, 2003.
- [132] M. C. Peckerar, "Reliable metallization for VLSI," in *Electromigration and Electronic Device Degradation* (A. Christou, ed.), pp. 317–338, John Wiley Sons, 1993.
- [133] E. Kinsbron, "A model for the width dependence of electromigration lifetimes in aluminum thin-film stripes," *Applied Physics Letters*, vol. 36, no. 12, pp. 968–970, 1980.
- [134] M. Scherge, V. Breternitz, and C. Knedlik, "Simulation of electromigration behavior in aluminum metallization of integrated circuits," *Microelectronics Reliability*, vol. 32, no. 1-2, pp. 21–24, 1992.
- [135] M. Shatzkes and J. R. Lloyd, "A model for conductor failure considering diffusion concurrently with electromigration resulting in a current exponent," *Journal of Applied Physics*, vol. 59, pp. 3890–3893, 1986.
- [136] J. R. Lloyd, "Reliability modeling for electromigration failure," *Quality and Reliability Engineering International*, vol. 10, pp. 303–308, 1994.
- [137] C. K. Hu, L. Gignac, and R. Rosenberg, "Electromigration of Cu/low dielectric constant interconnects," *Microelectronics Reliability*, vol. 46, pp. 213–231, 2006.
- [138] E. Glickman and M. Nathan, "On the unusual electromigration behavior of copper interconnects," *Journal of Applied Physics*, vol. 80, no. 7, pp. 3782–3791, 1996.
- [139] J. R. Lloyd, J. Clemens, and R. Snede, "Copper metallization reliability," *Microelectronics Reliability*, vol. 39, pp. 1595–1602, 1999.
- [140] B. N. Agarwala, M. J. Attardo, and A. P. Ingraham, "Dependence of electromigration-induced failure time on length and width of aluminum thin-film conductors," *Journal of Applied Physics*, vol. 41, no. 10, pp. 3954–3960, 1970.
- [141] T. Kwok, J. Finnegan, and D. Johnson, "Effects of linewidth and bend structure on electromigration lifetime in Al-Cu submicron interconnects," in *Fifth International IEEE VLSI Multilevel Interconnection Conference*, pp. 436–445, 1988.

- [142] T. Nogami, S. Oka, K. Naganuma, and T. Nakata, "Electromigration lifetime as a function of line length or step number," in *1992 IEEE IRPS*, pp. 366–372, 1992.
- [143] Y.-K. Cheng, C.-H. Tsai, C.-C. Teng, and S.-M. Kang, *Electrothermal Analysis of VLSI Systems*. Kluwer Academic Publishers, 2000.
- [144] A. Christou, *Electromigration and Electronic Device Degradation*, ch. 9, pp. 291–316. Wiley Interscience, 1994.
- [145] J. M. Towner and E. P. V. de Ven, "Aluminum electromigration under pulsed DC conditions," in *21st Annual Proceedings of Reliability Physics Symposium*, (Phoenix, AZ, USA), pp. 36–39, IEEE, 1983.
- [146] L. Brooke, "Pulsed current electromigration failure model," in *25th Annual Proceedings-Reliability Physics*, pp. 136–139, IEEE, 1987.
- [147] J. J. Clement, "Vacancy supersaturation model for electromigration failure under DC and pulsed DC stress," *Journal of Applied Physics*, vol. 71, no. 9, pp. 4264–4268, 1992.
- [148] J. Tao, N. W. Cheung, and C. Hu, "An electromigration failure model for interconnects under pulsed and bidirectional current stressing," *IEEE Transactions on Electron Devices*, vol. 41, pp. 539–545, 1994.
- [149] B.-K. Liew, N. W. Cheung, and C. Hu, "Projecting interconnect electromigration lifetime for arbitrary current waveforms," *IEEE Transactions on Electron Devices*, vol. 37, no. 5, pp. 1343–1351, 1990.
- [150] L. M. Ting, J. S. May, W. R. Hunter, and J. W. McPherson, "AC electromigration characterization and modeling of multilayered interconnects," in *IRPS*, pp. 300–316, 1993.
- [151] J. R. Black, "Current limitations of thin film conductors," in *IRPS*, pp. 300–306, 1982.
- [152] F. Chen, J. Gill, D. Harmon, T. Sullivan, B. Li, A. Strong, H. Rathore, D. Edelstein, C. C. Yang, A. Cowley, and L. Clevenger, "Measurements of effective thermal conductivity for advanced interconnect structures with various composite low-k dielectrics," in *IRPS*, pp. 68–73, 2004.
- [153] H. A. Schafft, "Thermal analysis of electromigration test structures," *IEEE Transactions on Electron Devices*, vol. ED-34, no. 3, pp. 664–672, 1987.
- [154] W. R. Hunter, "Self-consistent solutions for allowed interconnect current density-part I: implications for technology evolution," *IEEE Transactions on Electron Devices*, vol. 44, no. 2, pp. 304–309, 1997.

- [155] W. R. Hunter, "Self-consistent solutions for allowed interconnect current density-part II: application to design guidelines," *IEEE Transactions on Electron Devices*, vol. 44, no. 2, pp. 310–316, 1997.
- [156] D. Harmon, J. Gill, and T. Sullivan, "Thermal conductance of IC interconnects embedded in dielectrics," in *IRW*, pp. 1–9, 1998.
- [157] B. Li, D. Harmon, J. Gill, F. Chen, and T. Sullivan, "Thermal and electromigration challenges for advanced interconnects," in *IRW*, pp. 46–51, 2004.
- [158] J. Cho and C. Thompson, "Grain size dependence of electromigration-induced failures in narrow interconnects," *Applied Physics Letter*, vol. 54, pp. 2577–2579, 1989.
- [159] A. S. Oates, "Electromigration failure distribution of contacts and vias as a function of stress conditions in submicron IC metallizations," in *34th IEEE IRPS*, (Dallas, TX), pp. 164–171, 1996.
- [160] K. Hinode, T. Furusawa, and Y. Homma, "Dependence of electromigration lifetime on the square of current density," in *IEEE IRPS 1993*, pp. 317–326, 1993.
- [161] J. A. Schwarz, "Distributions of activation energies for electromigration damage in thin-film aluminum interconnects," *Journal of Applied Physics*, vol. 61, pp. 798–800, 1987.
- [162] J. R. Lloyd, "On the log-normal distribution of electromigration lifetimes," *Journal of Applied Physics*, vol. 50, pp. 5062–5064, 1979.
- [163] A. Bobbio and O. Saracco, "On the spread of time-to-failure measurements in thin metallic films," *Thin Solid Films*, vol. 17, no. 1, pp. S13–S16, 1973.
- [164] M. J. Attardo, R. Rutledge, and R. C. Jack, "Statistical metallurgical for electromigration failure in aluminum thin-film conductors," *Journal of Applied Physics*, vol. 42, pp. 4343–4349, 1971.
- [165] J. M. Towner, "Are electromigration failures lognormal distributed?," in *IEEE IRPS*, pp. 100–105, 1990.
- [166] M. Gall, C. Capasso, D. Jawarani, R. Hernandez, and H. Kawasaki, "Statistical analysis of early failures in electromigration," *Applied Physics*, vol. 90, no. 2, pp. 732–740, 2001.
- [167] C. Pennetta, L. Reggiani, G. Trefan, F. Fantini, A. Scorzoni, and I. D. Munari, "A percolative approach to electromigration in metallic lines," *Journal of Physics D: Applied Physics*, vol. 3, pp. 1421–1429, 2001.
- [168] B. Li, T. D. Sullivan, and T. C. Lee, "Line depletion electromigration characterization of Cu interconnects," *IEEE Transactions on Device and Materials Reliability*, vol. 4, pp. 80–85, 2004.

- [169] J. B. Lai, J. L. Yang, Y. P. Wang, S. H. Chang, R. L. Hwang, Y. S. Huang, and C. S. Hou, "A study of bimodal distributions of time-to-failure of copper via electromigration," in *International Symposium on VLSI Technology, Systems, and Applications*, pp. 271–274, 2001.
- [170] E. T. Ogawa, K. D. Lee, and H. Matsushashi, "Statistics of electromigration early failures in Cu/oxide dual-damascene interconnects," in *39th Annual International Reliability Physics Symposium*, (Orlando,Florida), pp. 341–349, 2001.
- [171] J. Gill, T. Sullivan, S. Yankee, H. Barth, and A. V. Glasow, "Investigation of via-dominated multi-model electromigration failure distributions in dual damascene Cu interconnects with a discussion of the statistical implication," in *40th IRPS*, (Daltas,Texas), pp. 298–304, 2002.
- [172] L. M. Gignac, C.-K. Hu, and E. G. Liniger, "Correlation of electromigration lifetime distribution to failure mode in dual damascene Cu/SiLK interconnects," *Microelectronic Engineering*, vol. 70, pp. 398–405, 2003.
- [173] M. H. Lin, Y. L. Lin, G. S. Yang, M. S. Yeh, K. P. Chang, K. C. Su, and T. Wang, "Comparison of copper interconnect electromigration behaviors in various structures for advanced BEOL technology," in *Proceedings of 11th IPFA*, pp. 177–180, IEEE, 2004.
- [174] K.-D. Lee and P. S. Ho, "Statistical study for electromigration reliability in dual-damascene Cu interconnects," *IEEE Transactions on Device and Materials Reliability*, vol. 4, no. 2, pp. 237–245, 2004.
- [175] R. G. Filippi, J. F. McGrath, T. M. Shaw, C. E. Murray, H. S. Rathore, P. S. McLaughlin, V. McGahay, L. Nicholson, P.-C. Wang, J. R. Lloyd, M. Lane, R. Rosenberg, X. Lin, Y. Y. Wang, W. Landers, T. Spooner, J. J. Demarest, B. H. Engel, J. Gill, G. Goth, E. Berth, G. Biery, C. R. Davis, R. A. Wachnik, R. Goldblatt, T. Ivers, A. Swinton, C. Barile, and J. Aitken, "Thermal cycle reliability of stacked via structures with copper metallization and an organic low-k dielectric," in *IRPS*, pp. 61–67, 2004.
- [176] C. J. Christiansen, B. Li, J. Gill, R. Filippi, and mathew Angyal, "Via-depletion electromigration in copper interconnects," *IEEE Transactions on Device and Materials Reliability*, vol. 6, no. 2, pp. 163–171, 2006.
- [177] A. H. Fischer, A. Abel, M. Lepper, A. E. Zitzelsberger, and A. V. Glasow, "Experimental data and statistical models for bimodal EM failures," in *38th IEEE IRPS*, (San Jose), pp. 359–364, 2000.
- [178] W. Denson, "The history of reliability prediction," *IEEE Transactions on Reliability*, vol. 47, no. 3-SP, pp. SP321–SP328, 1998.

- [179] B. Foucher, J. Boullie, B. Meslet, and D. Das, “A review of reliability prediction methods for electronic devices,” *Microelectronics Reliability*, vol. 42, pp. 1155–1162, 2002.
- [180] A. Haggag, M. Lemanski, G. Anderson, P. Abramowitz, and M. Moosa, “Realistic projections of product fmax shift and statistics due to HCI and NBTI,” in *IRPS*, pp. 93–96, 2007.
- [181] W. Bornstein, R. Dunn, and T. Spielberg, “Field degradation of memory components due to hot carriers,” in *IRPS*, pp. 294–298, 2006.
- [182] “Telcordia technologies special report sr-332,” tech. rep., Telcordia Technologies, Inc, 2001.
- [183] *Modele universel pour le calcul de la fiabilite previsionelle des composants, cartes et equipements electroniques*, 2000.
- [184] “Handbook of reliability data,” tech. rep., British Telecom, 1995.
- [185] “Siemens company standard sn29500, version 6.0, failure rates of electronic components,” tech. rep., Siemens AG, 1999.
- [186] “Standard reliability table for semiconductor devices,” tech. rep., Nippon Telegraph and Telephone Corporation, 1985.
- [187] D. D. Dylis and M. G. Priore, “A comprehensive reliability assessment tool for electronic systems,” in *RAMS*, pp. 308–313, 2001.
- [188] “FIDES guide 2004 issue A reliability methodology for electronic systems,” tech. rep., FIDES Group, 2004.
- [189] M. G. Pecht and F. R. Nash, “Predicting the reliability of electronic equipment,” *Proceedings of the IEEE*, vol. 82, pp. 992–1004, July 1994.
- [190] M. Economou, “The merits and limitations of reliability predictions,” in *RAMS*, pp. 352–357, 2004.
- [191] Q. K. Zhu, *Power Distribution Network Design for VLSI*. Wiley Interscience, 2004.
- [192] I. A. Blech, “Electromigration in thin aluminum films on titanium nitride,” *Journal of Applied Physics*, vol. 47, pp. 1203–1208, 1976.
- [193] M. White, D. Vu, D. Nguyen, R. Ruiz, Y. Chen, and J. B. Bernstein, “Product reliability trends, derating considerations and failure mechanisms with scaled CMOS,” in *IIRW*, pp. 156–159, 2006.
- [194] M. Ershov, S. Saxena, S. Minehane, P. Clifton, M. Redford, R. Lindley, H. Karbasi, S. Graves, and S. Winters, “Degradation dynamics, recovery, and characterization of negative bias temperature instability,” *Microelectronics Reliability*, vol. 45, pp. 99–105, 2005.

- [195] G. L. Rosa, F. Guarin, S. Rauch, A. Acovic, J. Lukaitis, and E. Crabbe, "NBTI-channel hot carrier effects in pmosfets in advanced CMOS technologies," in *IRPS*, pp. 282–286, IEEE, 1997.
- [196] S. H. Gunther, F. Binns, D. M. Carmean, and J. C. Hall, "Managing the impact of increasing microprocessor power consumption," *Intel Technology Journal*, vol. Q1, pp. 1–9, 2001.
- [197] K. Skadron, M. R. Stan, W. Huang, and S. Velusamy, "Temperature-aware microarchitecture," in *30th International Symposium on Computer Architecture*, pp. 2–13, 2003.
- [198] H. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*. Addison-Wesley Company, 1990.
- [199] J. P. Uyemura, *Introduction to VLSI Circuits and Systems*. John Wiley & Sons, INC, 2002.
- [200] N. Weste and D. Harris, *CMOS VLSI design: a circuits and systems perspective*. Addison Wesley, 2004.
- [201] ARM, *Advantage SRAM Generator User Guide*, 2006.
- [202] IBM, *IBM CMOS 9SF Technology Design Manual*, 2006.
- [203] X. Li, J. Qin, B. Huang, X. Zhang, and J. Bernstein, "SRAM circuit-failure modeling and reliabilitysimulation with SPICE," *IEEE Transactions on Device and Materials Reliability*, vol. 2, pp. 235–246, 2006.
- [204] D. Edelstein, H. Rathore, C. Davis, L. Clevenger, A. Cowley, T. Nogami, B. Agarwala, S. Arai, A. Carbone, K. Chanda, F. Chen, S. Cohen, W. Cote, M. Cullinan, T. Dalton, S. Das, P. Davis, J. Demarest, D. Dunn, and C. Dziobkowski, "Comprehensive reliability evaluation of a 90 nm CMOS technology with Cu/PECVD low-k BEOL," in *IRPS*, pp. 316–319, 2004.