

## ABSTRACT

Title of dissertation: DEEP SUBMICRON CMOS VLSI CIRCUIT  
RELIABILITY MODELING, SIMULATION  
AND DESIGN

Xiaojun Li, Doctor of Philosophy, 2005

Dissertation directed by: Professor Joseph B. Bernstein  
Reliability Engineering

CMOS VLSI circuit reliability modeling and simulation have attracted intense research interest in the last two decades, and as a result almost all IC Design For Reliability (DFR) tools now try to incrementally simulate device wearout mechanisms in iterative ways. These DFR tools are capable of accurately characterizing the device wearout process and predicting its impact on circuit performance. Nevertheless, excessive simulation time and tedious parameter testing process often limit popularity of these tools in product design and fabrication.

This work develops a new SPICE reliability simulation method that shifts the focus of reliability analysis from device wearout to circuit functionality. A set of accelerated lifetime models and failure equivalent circuit models are proposed for the most common MOSFET intrinsic wearout mechanisms, including Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), and Negative Bias Temperature Instability (NBTI). The accelerated lifetime models help to identify the most degraded transistors in a circuit in terms of the device's terminal voltage

and current waveforms. Then corresponding failure equivalent circuit models are incorporated into the circuit to substitute these identified transistors. Finally, SPICE simulation is performed again to check circuit functionality and analyze the impact of device wearout on circuit operation. Device wearout effects are lumped into a very limited number of failure equivalent circuit model parameters, and circuit performance degradation and functionality are determined by the magnitude of these parameters.

In this new method, it is unnecessary to perform a large number of small-step SPICE simulation iterations. Therefore, simulation time is obviously shortened in comparison to other tools. In addition, a reduced set of failure equivalent circuit model parameters, rather than a large number of device SPICE model parameters, need to be accurately characterized at each interim wearout process. Thus device testing and parameter extraction work are also significantly simplified. These advantages will allow circuit designers to perform quick and efficient circuit reliability analyses and to develop practical guidelines for reliable electronic designs.

DEEP SUBMICRON CMOS VLSI CIRCUIT  
RELIABILITY MODELING, SIMULATION AND DESIGN

by

Xiaojun Li

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  
2005

Advisory Committee:

Professor Joseph B. Bernstein, Chair/Advisor  
Professor Aris Christou  
Professor Ali Mosleh  
Professor Martin Peckerar  
Professor Carol Smidts

## DEDICATION

I would like to dedicate this work to my pretty daughter Kemeng, who entered the world while I was working on this dissertation. Her first cry delivers to me the great pleasure of creation and being fatherhood. Her tender yet beautiful life sprouts like a bud in early spring, bearing vital force and hope, which always inspires me to work diligently and strive toward excellence.

## ACKNOWLEDGMENTS

I owe my gratitude to all the professors and colleagues who have made contributions to this dissertation. Their company and persisting guidance make my study and research experience of the last three years in the University of Maryland a period of rewarding time in my life journey.

First of all, I would like to thank my advisor, Professor Joseph B. Bernstein, for providing financial support for my study since I came to Maryland in summer 2002. He also granted me privileges to work on various challenging and interesting projects in the field of microelectronics reliability. With his sharp insight, he laid the foundation and formulated some of the original ideas for this dissertation work which perfectly matches my expertise and interest. He is one of the most gracious and knowledgeable professors I ever met in my life. It is a great pleasure to work with him and learn from him both in academic and in everyday life during the last several years.

I would also like to thank those professors serving on my dissertation committee. I became acquainted with them either from taking their courses or participating in their projects. Their profound knowledge and enlightening discussions instilled many essential ingredients into this dissertation work. Their persisting support and invaluable guidance are proved to be indispensable for me to improve my research capability.

Furthermore, I sincerely appreciate Dr. Ernest Wu at IBM Microelectronics Division and Dr. Sufi Zafar at IBM T. J. Watson Research Center. They are very generous to share pre-published papers with me. Without their help, I could not finish the modeling work of some wearout mechanisms in a short time.

Finally, I owe my deepest gratitude to my wife. She gave up her successful career and comfortable life in Shanghai and came to Maryland to support my study. We have been staying together and living through the hard time when I was pursuing the doctoral degree. She is the most dependable partner whenever I head for a new higher life goal.

# TABLE OF CONTENTS

List of Figures	viii
1 Introduction	1
1.1 CMOS Scaling and New Reliability Challenges . . . . .	1
1.2 Purpose of the Dissertation . . . . .	4
1.3 Dissertation Organization and Chapter Overviews . . . . .	4
2 Simulation Models and Algorithms	10
2.1 Overview . . . . .	10
2.2 Review of Reliability Simulation Tools . . . . .	12
2.2.1 Reliability Simulation in Virtuoso UltraSim . . . . .	12
2.2.2 Reliability Simulation in Eldo . . . . .	14
2.3 Limitations and Improvements . . . . .	17
2.4 Assumptions and Justification . . . . .	22
2.5 Summary of Lifetime and Circuit Models . . . . .	24
2.5.1 Hot Carrier Injection . . . . .	25
2.5.2 Time Dependent Dielectric Breakdown . . . . .	25
2.5.3 Negative Bias Temperature Instability . . . . .	27
2.6 Reliability Prediction and Simulation Algorithms . . . . .	29
2.6.1 Circuit Lifetime and Failure Rate Prediction . . . . .	29
2.6.2 Circuit Reliability Simulation Algorithm . . . . .	33
3 Hot Carrier Injection Effect and Models	36
3.1 Introduction . . . . .	36
3.2 Accelerated Lifetime Model . . . . .	37
3.3 Failure Equivalent Circuit Model . . . . .	43
3.4 Implementation in MaCRO . . . . .	52
3.5 Conclusion . . . . .	53
4 Time Dependent Dielectric Breakdown Effect and Models	54
4.1 Introduction . . . . .	54
4.2 Accelerated Lifetime Model . . . . .	55
4.3 Failure Equivalent Circuit Model . . . . .	62
4.4 Implementation in MaCRO . . . . .	76
4.5 Conclusion . . . . .	79
5 Negative Bias Temperature Instability Effect and Models	80
5.1 Introduction . . . . .	80
5.2 Accelerated Lifetime Model . . . . .	81
5.3 Failure Equivalent Circuit Model . . . . .	90
5.4 Implementation in MaCRO . . . . .	94
5.5 Conclusion . . . . .	96

6	Electromigration Lifetime Models and Parameter Extraction	98
6.1	Introduction	98
6.2	Electromigration Failure Physics	99
6.2.1	Nucleation-Dominated Failure Physics	100
6.2.2	Growth-Dominated Failure Physics	100
6.3	Electromigration Lifetime Models	101
6.4	Model Parameter Extraction	102
6.4.1	Current Density	103
6.4.2	Current Acceleration Factor and Activation Energy	105
6.4.3	Temperature Effects	108
6.4.4	Example of Typical Values	110
6.5	Electromigration of Copper Wires	110
7	Derating Voltage and Temperature for Reliability	112
7.1	Introduction	112
7.2	Circuit Design and Simulation	114
7.3	Simulation Results and Analysis	115
7.3.1	Voltage Derating Analysis	115
7.3.2	Temperature Derating Analysis	118
7.3.3	Voltage Transfer Analysis	122
7.4	Derating Model and Derating Factor	125
7.5	Derating Factor and Simulation	128
7.6	Conclusion	130
8	SRAM Reliability Simulation and Analysis	132
8.1	Introduction	132
8.2	SRAM Circuit Design and Simulation	133
8.3	Preview of SRAM Failure Behaviors	138
8.4	Device Lifetime Calculation	141
8.5	SPICE Reliability Simulation with Circuit Models	145
8.5.1	HCI	146
8.5.2	HCI+TDDB	149
8.5.3	HCI+TDDB+NBTI	152
8.6	Reliability Design Techniques	157
8.7	Conclusion	158
9	Summary	159
9.1	Results	159
9.2	Main Contributions	162
9.3	Future Work	163
A	MATLAB Programs for Circuit Model Calculation	165
A.1	Hot Carrier Injection	165
A.2	Time Dependent Dielectric Breakdown	166
A.3	Negative Bias Temperature Instability	168

B	Flowchart of Lifetime and Reliability Trend Prediction	169
C	Terms and Abbreviations	170
	Bibliography	172

## LIST OF FIGURES

2.1	Hot carrier reliability simulation flowchart in Virtuoso UltraSim. Device wearout modeling is the focus of the reliability analysis [17]. . . .	13
2.2	HCI reliability simulation in Eldo [19]. A large number of SPICE simulation iterations have to be carried out to obtain accuracy. . . .	16
2.3	HCI circuit model in MaCRO. In the model: $V_{gdx} = V_{gs} - V_t - V_{ds}$ and $V_{Rd} = I_{ds}\Delta R_d$ . $V_t$ is threshold voltage and $I_{ds}$ is the current from node $D$ to $S$ . . . . .	26
2.4	TDDB circuit model in MaCRO. $I_{OX} = I_S - I_D$ is a voltage-dependent current source representing breakdown path current injection effect. $R_D$ and $R_S$ characterize the resistance in the source and the drain extensions, respectively. $L_1$ represents breakdown location in terms of the source edge. . . . .	27
2.5	NBTI circuit model in MaCRO. The inclusion of $I_{GD}$ and $I_{GS}$ inherently accounts for oxide breakdown effects and also supplies leakage currents for $R_G$ whose voltage drop is equivalent to pMOSFET threshold voltage degradation. . . . .	28
2.6	Flowchart of device and circuit lifetime and failure rate prediction process with MaCRO lifetime models. SPICE simulation predicts device terminal voltage and current stress profiles, and model fitting parameters are determined from device testing work. . . . .	32
2.7	MaCRO circuit reliability simulation algorithm. . . . .	35
3.1	BERT nMOSFET HCI circuit model. (a) Bidirectional interface trap generation near both drain and source. $L_f$ and $L_r$ represent forward and reverse hot carrier damaged regions. (b) HCI drain current $\Delta I_d$ circuit model [52]. . . . .	45
3.2	UIUC nMOSFET HCI two-transistor series model. (a) Triangular oxide charge distribution profile used in model derivation. (b) Cross-sectional view of nMOSFET with hot carrier damage, $L_2$ is damaged channel region. (c) Two-transistor series circuit model. The parasitic transistor has different channel mobility and threshold voltage with the channel length $L_2$ set to $0.1\mu m$ [46, 51, 55]. . . . .	46
3.3	HCI circuit model in MaCRO. In the model: $V_{gdx} = V_{gs} - V_t - V_{ds}$ and $V_{Rd} = I_{ds}\Delta R_d$ . $V_t$ is threshold voltage and $I_{ds}$ is the current from node $D$ to $S$ . . . . .	48

4.1	Power-law leakage current model. The exponent $p$ varies from 5 to 2 as the degradation level increases. $K$ reflects the “size” of the breakdown spot. . . . .	64
4.2	TDDDB GOS model for gate-to-channel breakdown of nMOSFET with $n^+$ -poly gate. The channel lengths of nMOS1 and nMOS2 follow the relation: $L_1 + L_2 = L$ where $L$ is the undamaged nMOSFET channel length. The parameter $R_{GOS}$ is related to the size and location of the breakdown path. A value of $R_{GOS}$ as low as $3K\Omega$ was used in the simulation in [82]. . . . .	65
4.3	TDDDB RF equivalent circuit model. Model parameters for simulation in [84] are set as: $R_G = 85.4\Omega$ , $R_D = R_S = 12.14\Omega$ , $R_{GD} = 6.88K\Omega$ , $R_{GS} = 23K\Omega$ , $C_{GDO} = C_{GSO} = 15.3fF$ , $C_{jDB} = C_{jSB} = 7fF$ , $R_{DSB} = 80K\Omega$ , $R_{DB} = R_{SB} = 49.37\Omega$ . . . . .	67
4.4	TDDDB circuit model for nMOSFET with hard gate oxide breakdown and operated in positive gate voltage. (a) Cross-sectional view of breakdown structure. (b) Equivalent circuit model. Model parameters for simulation in [86] are set as: $R_G = 1K\Omega$ , $L_{MS} + L_{MS} = 0.09\mu m$ , $W_{MS} = W_{MS} = 0.25\mu m$ , $R_D$ and $R_S$ vary from $2.5K\Omega$ (at source and drain) to $12.5K\Omega$ (at the middle of the channel). . . . .	70
4.5	MaCRO TDDDB circuit model for nMOSFET with hard gate oxide breakdown. $I_{OX} = I_S - I_D$ is a voltage dependent current source representing breakdown path current injection effect. $R_D$ and $R_S$ characterize the resistance in the source and the drain extensions, respectively. $L_1$ represents breakdown location away from the source edge. . . . .	74
4.6	Relationship between effective gate resistance $R_G$ and oxide breakdown location $L_1$ . The bias condition is $V_G = 1.5V$ , $V_D = V_S = 0$ . (a) Reproduce from [87] with illustrative data. (b) Cross-sectional view of breakdown location. . . . .	77
5.1	The different stress time windows of HCI and NBTI for an inverter in (a) VTC plot and (b) input-output waveform plot. HCI stresses devices only during the dynamic switching periods when both gate voltage and drain voltage are high enough and there is current flowing through the device. NBTI stresses pMOS devices mainly during the period when they are in one of the two static operation states when gate voltage is negative with respect to drain and source voltages. . .	82

5.2	MaCRO NBTI circuit model. NBTI-induced pMOSFET threshold voltage increase is modeled as absolute gate-to-source voltage decrease. Gate tunneling current flowing through the gate resistance $R_G$ leads to the increase of voltage at point $G'$ . This corresponds to the decrease of pMOSFET absolute gate-to-source voltage and therefore mimics the threshold voltage degradation effect. Gate tunneling current is modeled with two voltage controlled current sources which follow the form of a power law relation as: $I = KV^P$ . . . . .	92
6.1	Plot of $\ln MTTF$ vs. $\ln J$ to illustrate the fitting method in extracting $n$ and judge the linearity of the dependence [138]. . . . .	107
6.2	Plot of $\ln MTTF$ vs. $1/T$ to illustrate the fitting method in extracting $E_{aEM}$ and judge the linearity of the dependence [138]. . . . .	108
7.1	The schematic diagram of the ring oscillator which consists of 17-stage CMOS inverters and interconnecting capacitors. . . . .	115
7.2	The derating relationship of frequency and power dissipation vs. $V_{DD}$ . When voltage increases 4 times, the frequency increases about 5 times, whereas the power dissipation increases near to 100 times. . . .	117
7.3	The derating curves of frequency and power dissipation vs. temperature. Operation frequency and power dissipation follow nonlinear trends when temperature is derated and simulation shows a minimum frequency value at temperature around $120^\circ C$ . . . . .	120
7.4	The simulation results for $NM_L$ and $NM_H$ vs. $V_{DD}$ . Over the voltage derating range of $4.0V$ to $1.2V$ , $NM_L$ and $NM_H$ approximately decrease linearly with $V_{DD}$ . . . . .	123
7.5	DC VTC curves at different power supply voltage (from $0.5V$ to $4.5V$ ). When $V_{DD}$ scales down, the transition region of VTC reduces proportionally. When $V_{DD}$ is lower than $0.8V$ , the transition region disappears, and the VTC exhibits a hysteresis behavior. . . . .	124
7.6	Trends of $D_f$ vs. $V_{DD}/V_{DD}^0$ with typical model parameters from the $0.18\mu m$ technology. $V_{DD}$ is derated within the range [100% ~ 80%] of its rated value $V_{DD}^0 = 3.3V$ . The trend of $D_f$ when $V_{DD}^0 = 1.8V$ is also plotted for comparison. . . . .	130
8.1	The one bit SRAM structural block diagram. The circuit consists of one bit 6-T cell, read/write control logic and output sense amplifier. . . . .	133

8.2	Schematic of the one bit 6-T SRAM cell. Store/Storen represent cell state. WORD line enables the two pass transistors $M5$ and $M6$ during memory read and write cycles. . . . .	135
8.3	Schematic of the precharge circuit. BIT/BITn lines are pre-charged to the same voltage level before each read and write operation. $M26 \sim M29$ are included for reducing transient charge injection effects. . . .	136
8.4	Schematic of the write control circuit. WRITE signal controls the connections of DATA line and BITio/BITnio lines. BITio/BITnio lines are connected to BIT/BITn by the two transmission gates ( $M41 \sim M44$ ). . . . .	137
8.5	Schematic of the sense amplifier. READ signal controls the operation of the latch amplifier and the connection between BIT/BITn and the output. The latch amplifier magnifies BIT/BITn line swings to full digital levels. . . . .	138
8.6	SRAM SPICE simulation stimuli. PRE exerts before each read/write operation. CD signal enables the transmission gates $M41 \sim M42$ and WORD signal enables the pass transistors $M5 \sim M6$ during each read/write operation. The “0” or “1” is available on DATA line during each write operation. . . . .	139
8.7	SRAM SPICE simulation results. (a) shows waveforms of BIT/BITn signals, (b) shows SRAM cell state signals Store/Storen, (c) is write operation result, and (d) is read operation result. . . . .	140
8.8	Device lifetime calculation results for the three failure mechanisms: (a) HCI, (b) TDDB, and (c) NBTI. The horizontal axis denotes device’s index, and vertical axis denotes lifetime value normalized to $\tau_1 \sim \tau_3$ , respectively. . . . .	143
8.9	Comparison of transition delay of $M58$ before and after inverter sizing. Proper sizing significantly reduces dynamic switching delay, thereby suppressing HCI effect. $W_n = 0.6\mu m$ before sizing, and $W_n = 1.8\mu m$ after sizing. . . . .	144
8.10	$\Delta R_d$ values of $M5, M6, M52$ at different stress times. The unit of horizontal axis is time in years, the vertical axis is in logarithmic scale and in unit Ohm. . . . .	147
8.11	The simulated waveforms of the SRAM cell Store signal and output DATAO signal after different stress times. At $t = 1$ year, Store signal does not jump to high as expected during the “write 1” cycle indicating failure of SRAM cell. . . . .	148

8.12	The waveforms of SRAM Store/Storen signals and BIT/BITn signals before and after circuit failure. Store/Storen signals do not flip due to the degradation in BIT/BITn signals when a reverse value is being written to the SRAM cell. . . . .	149
8.13	The simulated waveforms of the SRAM cell Store signal and output DATAO signal at different HCI+TDDB stress times. At $t = 0.6$ year, Store signal does not jump to high during the “write 1” cycle indicating failure of SRAM cell. . . . .	150
8.14	The waveforms of the SRAM Store/Store signals and BIT/BITn signals before and after circuit failure. . . . .	151
8.15	The TDDB effect of $M53$ on sense amplifier output stability. The breakdown in $M53$ provides additional current path between BITn <sub>io</sub> and $V_{DD}$ and tends to pull up BITn <sub>io</sub> when it is at low level in “read 0” and “write 1” cycles. . . . .	152
8.16	The TDDB+NBTI circuit model for a pMOSFET. $R_G$ and $I_{ox}$ account for threshold voltage degradation due to NBTI. $I_{ox}$ and the two split pMOSFETs represent TDDB damage. $R_D$ and $R_S$ characterize the resistances in drain and source extensions. They are excluded in this SRAM case study in order to simplify simulation work. . . . .	153
8.17	The NBTI effects on SRAM cell transition speed. The switching speed of SRAM cell Store/Storen signals degrades when NBTI damage on $M3$ is considered. . . . .	154
8.18	Voltage transfer curves of BITn-to-Storen for different combinations of failure mechanisms. From left to right, the curves represent effects of TDDB, no damage, HCI+TDDB+NBTI, HCI, and HCI+NBTI, respectively. NBTI has negligible effect on SRAM cell stability. . . . .	155
8.19	Butterfly plots for various failure mechanisms. (a) denotes the no-damage operation, (b) shows SNM degradation due to TDDB, (c) shows the combined effect of TDDB+NBTI, and (d) is the combination of the previous three plots. The difference in (b) and (c) is very small indicating that NBTI is not a dominant effect. . . . .	156
B.1	MaCRO Flow of lifetime, failure rate and reliability trend prediction.	169

# Chapter 1

## Introduction

### 1.1 CMOS Scaling and New Reliability Challenges

The scaling of CMOS technology into deep submicron regimes has brought about new reliability challenges, which are forcing dramatic changes in approaches to integrated circuit reliability assurance. Product cost and performance requirements will be substantially affected, or even superseded, by reliability constraints [1]. The traditional reliability assurance methods, which relied on failure detection and analysis at the end of a lengthy product development process, are rapidly losing efficiency due to the reliability trends predicted by 2003 International Technology Roadmap for Semiconductor (ITRS'03) [2].

For most applications, current overall chip reliability levels need to be maintained over the next fifteen years, despite the possible risks induced by multiple major technology breakthroughs. This constraint requires continuous improvement in reliability per transistor and per unit length of metal interconnect due to the continuous shrinkage of device dimensions. Scaling pushes device performance to the limits of technology and gradually eats up circuit reliability margins. Therefore, the accurate tradeoffs between performance and reliability must be addressed before

committing design to production.

The projected failure in time (FIT) of technology nodes from  $90nm$  to  $65nm$  in ITRS'03 is on the order of 10 to 100. However, experimentally determining FIT values this low by traditional reliability qualification methods requires a huge number of device-hours of Accelerated Life Testing (ALT). Approximately  $9 \times 10^7$  device-hours of testing are required to prove a failure rate of 10 FITs at 60% confidence level if no failures occur during the testing [3]. The increased cost and excessive time consumed by testing work demand that accurate lifetime models and efficient reliability simulation tools must be available in product design stages.

The validity of the voltage and temperature acceleration methods that have been utilized in reliability screening and qualification processes, such as burn-in and ALT, becomes questionable due to the diminished margins for proper acceleration of these stress factors. The traditional FIT and acceleration factor determination methods that rely on the multiplication of individual acceleration factors need to be revisited, and the correlation of these factors must be explored and modeled for the purpose of accurate failure rate prediction.

Finally, as circuits become increasingly complex, two irreversible trends can be noted: First, a given device within a chip is stressed for a decreasing fraction of the reliability testing time; Second, a longer delay is required to correct the reliability problem by process and design iterations [4].

All of the above trends demand that device lifetime and circuit reliability be accurately characterized and predicted during the product design process. This can only be fulfilled by effective IC reliability simulation tools.

CMOS circuit reliability simulation has attracted intense research interest in the last two decades. Significant progress in modeling device wearout mechanisms has led to the emergence of quite a few successful reliability simulation tools [5, 6, 7, 8, 9]. The simulation algorithms adopted by these tools physically characterize the device wearout process under real circuit stress environments and incrementally simulate circuit performance degradation in iterative ways. This physics-of-failure based iterative simulation algorithm often produces accurate simulation results with the disadvantage of excessive computational and experimental work. Some attempts have been made to improve simulation efficiency by employing the fast timing simulation method [10, 11] or by performing gate-level circuit simulation [12]. However, the device wearout-based simulation and testing philosophy is preserved. As a result, even though reliability simulation is generally regarded as an essential step in deep submicron CMOS circuit designs, the tedious device aging test and model parameter extraction work often discourage chip designers from exercising IC reliability simulation in their everyday work.

In review of reliability simulation practice in industrial and academic communities, it is obvious that some fundamental concepts and techniques have been universally adopted that not only form the common foundation of legacy reliability simulation tools but also nurture new ideas in some previously unresearched areas. These new ideas will give rise to developments and breakthroughs of new IC reliability simulation methods, which are both efficient and effective.

## 1.2 Purpose of the Dissertation

This dissertation focuses on developing a new *Maryland Circuit Reliability – Oriented* (MaCRO) SPICE simulation method, which is built upon the constant failure rate concept and equivalent circuit modeling techniques. MaCRO consists of a series of accelerated lifetime models and failure equivalent circuit models for common silicon intrinsic wearout mechanisms, including Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), and Negative Bias Temperature Instability (NBTI), respectively. In this new method, the overall simulation flow is straightforward, and SPICE engine is only initiated for very limited times to simulate the impact of device wearout on circuit functionality. Therefore, simulation time is obviously shortened. Also, a reduced set of failure equivalent circuit model parameters at each interim wearout process, rather than a large number of device SPICE model parameters, need to be accurately characterized. Thus, device testing and parameter extraction processes are also significantly simplified. These advantages allow circuit designers to perform quick and efficient circuit reliability analyses and to develop practical guidelines for reliable electronic designs.

## 1.3 Dissertation Organization and Chapter Overviews

This dissertation is organized into nine chapters, moving from MaCRO simulation algorithms to model developments to application examples. The overall structure follows a top-down-then-bottom-up presentation style: The first two chapters are a top-level overview of MaCRO simulation method and some other state-of-

the-art DFR tools; The middle four chapters present detailed developments of all MaCRO models and equations, which are basic ingredients of the MaCRO simulation algorithms; The following two chapters discuss two different kinds of top-level applications of MaCRO models by circuit simulation examples; Finally, the last chapter concludes the dissertation with a summary of contributions of this dissertation and suggestions for future work.

A brief overview of each chapter is given below to quickly walk readers through the overall dissertation. Note that for brevity, starting from the following paragraph, the term “accelerated lifetime model” is abbreviated to “lifetime model”, and the term “failure equivalent circuit model” is abbreviated to “circuit model”.

Chapter 2 provides readers a taste of the primary MaCRO models as well as overall simulation algorithms. First, two commercial state-of-the-art reliability simulation tools are reviewed, followed by a discussion of their limitations and possible improvements. Then, a set of MaCRO lifetime models and circuit models for each wearout mechanism are summarized. Finally, the overall MaCRO simulation algorithms, tailored for two distinct application purposes, are presented.

Chapters 3 to 5 introduce the detailed development of the lifetime and circuit models for HCI, TDDB and NBTI, respectively. Chapter 3 is dedicated to HCI effect and discusses the modeling process for this somewhat “old” wearout mechanism. In this chapter, the  $\Delta R_d$  model proposed elsewhere is improved to include the contributions of both interface trap generation and oxide charge trapping effects, the latter one being neglected in the original  $\Delta R_d$  model. This improvement, although complicating parameter extraction work, is physically more comprehensive

and accurate in characterizing hot carrier damages.

Chapter 4 focuses on TDDB effect and presents detailed development of lifetime and failure circuit models for this important wearout mechanism. This chapter proposes an advanced TDDB lifetime model that combines many important experimental observations, including power law voltage acceleration, non-Arrhenius temperature acceleration, Poisson area scaling statistics, and cumulative failure percentile scaling effects. Before introducing the new MaCRO circuit model, a thorough review of existing TDDB circuit models is presented for the sake of compensating for the obvious absence of overview papers of this kind in this area. From this careful review, an important error in the most frequently used TDDB SPICE circuit model is identified. Finally, a new TDDB circuit model is proposed and the number of model parameters is reduced to only one, which significantly simplifies its application process in circuit reliability analysis.

Chapter 5 covers NBTI effect and introduces lifetime and circuit models for this relatively new wearout mechanism. Based on an existing physics and statistics based model, a new NBTI lifetime model is developed that explains most experimental observations on NBTI-induced threshold voltage variations, including fractional power law dependence, saturation phenomenon, and dynamic recovery effects. Weibull statistics is included in explaining this NBTI lifetime model, thereby providing a new understanding of NBTI degradation behaviors. Starting from this new lifetime model, a physics-of-failure based NBTI circuit model, which is both simple and expandable, is developed. It is presumed to be the first NBTI damaged circuit model in literature.

Besides the aforementioned MOSFET-related wearout mechanisms, another important failure type intrinsic to Silicon chips is Electromigration (EM). Distinct from other mechanisms, EM is a metallization-related wearout process. For simplicity, MaCRO does not consider the impact of EM on circuit functionality. However, it includes the contribution of EM in circuit failure rate prediction and product derating behavior analysis. Chapter 6 focuses on EM failure physics and lifetime modeling, and provides practical guidelines in extrapolation of current density and activation energy to estimate EM failure rates. The EM models are integrated into MaCRO and combined with other lifetime models to help designers properly derate device and circuit operating parameters for reliability improvement and to predict reliability trends in future technologies.

The MaCRO models can be used for various application purposes depending on different data availability. If all lifetime model parameters can be obtained from experimental work, then MaCRO can accurately calculate the circuit lifetimes and failure rates. If process parameters of future technologies are projected, then MaCRO can predict reliability trends over generations and identify critical failure mechanisms. Based on the previous two applications, MaCRO can be further used in derating product voltage and temperature for reliability enhancement. If circuit functionality is of primary interest, MaCRO can quickly identify more damaged transistors in circuit in terms of the device's terminal voltage and current stress profiles, then MaCRO can include corresponding circuit models in the second round of SPICE simulation, which will reveal whether or not circuit functionality is maintained. Among these different applications, derating for reliability and circuit

reliability simulation are of primary importance and more difficult to implement. Chapter 7 addresses the derating method while chapter 8 shows a circuit example for reliability simulation with failure models.

Chapter 7 explores how device and circuit operating parameters, such as switching speed and power dissipation, scale with voltage and temperature. A 17-stage CMOS ring oscillator is simulated under different stress conditions to characterize the accurate derating relations and trends. Reduced voltage, frequency and temperature will reduce internal stresses in devices, thereby improving the devices' reliability. Since all these variations for a single device are proportional, the ratios can be applied to a full circuit with the help of a simple derating model. From the ring oscillator simulation, some practical design guidelines are formulated for developers to correctly derate devices for long-life applications.

In Chapter 8, a simple SRAM circuit is designed and simulated to demonstrate how to apply MaCRO to circuit reliability modeling, simulation, analysis and design. The SRAM circuit, implemented with a commercial  $0.25\mu\text{m}$  technology, consists of functional blocks of one bit 6-transistor (6-T) cell, precharge, read/write control and sense amplifier. The SRAM operation sequence of “write 0, read 0, write 1, read 1” is first simulated in SPICE to obtain terminal voltage and current stress profiles of each transistor. Then, normalized lifetimes of all transistors, in terms of each wearout mechanism, are calculated with the corresponding lifetime models. These lifetime values are sorted to single out the most damaged transistors. Finally, the selected transistors are substituted with circuit models, and SPICE simulation is performed again to characterize circuit performance, functionality, and

failure behaviors. Simulation shows that for the  $0.25\mu m$  technology, HCI and TDDB have significant effects on SRAM cell stability and Voltage Transfer Characteristics (VTC) while NBTI mainly degrades cell transition speed when the cell state flips. The illustrative SRAM simulation work proves by using MaCRO models that circuit designers can better understand the damage effects of HCI/TDDB/NBTI on circuit operation, quickly estimate circuit functional lifetime, make appropriate performance and reliability tradeoffs, and formulate practical design guidelines to improve circuit resistance to failures.

Chapter 9 concludes this dissertation. The main contributions of this work are summarized and some suggestions for future work are proposed.

## Chapter 2

### Simulation Models and Algorithms

#### 2.1 Overview

The advent of deep submicron technologies and the continuing shrinkage of MOSFET physical geometries have raised many new challenges in predicting circuit lifetimes and securing sufficient reliability margins. One of the essentials of reliable IC production is consistently fabricating a product that is capable of sustaining its intended functionality for specified time under stated operating conditions. The established practice has been to incrementally improve reliability through a lengthy design-manufacture-test cycle, however, this method is proved to be prohibitively expensive in most small-volume productions. Furthermore, some of the long-term wearout mechanisms cannot be identified and properly weeded out only by accelerated burn-in tests [13]. These considerations, in addition to fierce competition and higher pressure in achieving a shorter time-to-market objective, have impelled product reliability analysis to be addressed in advance at the initial design stage.

The development and use of effective reliability simulation methods are one of promising solutions for this early analysis and assessment. Once the reliability factors of the circuit are calibrated through simulation, the results can be compared

with initial specifications or limits. If the predicted reliability falls short of the requirements, new design iterations will be performed to improve the circuit robustness to failures. This cycle may be repeated for several times until the simulated reliability is satisfied. The ultimate goal of circuit reliability simulation, i.e. DFR, is toward Built-In-Reliability (BIR) allowing designers properly weigh performance and reliability tradeoffs and fully explore potentials of deep submicron technologies.

There are three distinct design levels at which DFR strategies are applicable: (a) at technology-level, where various material and structural failure mechanisms can be simulated with Technology Computer Aided Design (TCAD) tools; (b) at circuit-level, where the reliability of overall circuits and the impact of failure mechanisms on circuit operations can be characterized; (c) at package-level, where circuits as one entity are stressed mechanically, electrically and thermally for reliability qualification [14]. The focus of this dissertation work is on the circuit-level DFR implementation.

There are very few simulation tools built from scratch, and MaCRO is unexceptional. The best way to understand the similarity and differentiation between MaCRO and other simulation methods is reviewing them side by side in the same framework. In this chapter, first, two commercial state-of-the-art reliability simulation tools are reviewed. What follows is a discussion of their limitations and possible improvements. Finally, a snapshot of the MaCRO models and simulation algorithms is given. The subsequent chapters zoom into each of these models and present their development and applications in details.

## 2.2 Review of Reliability Simulation Tools

Hot carrier induced MOS device wearout is one of the most critical reliability issues for deep submicron CMOS integrated circuits. Hot carrier reliability models and simulation methods have been proposed and widely implemented in the semiconductor industry for many years. To some extent, the accuracy of hot carrier reliability simulation represents the robustness and efficiency of an entire reliability simulator, therefore, for simplicity, HCI simulation is employed here as the vehicle to deliver the basic concepts, modeling techniques and simulation flows realized in some commercial state-of-the-art reliability tools.

### 2.2.1 Reliability Simulation in Virtuoso UltraSim

Virtuoso UltraSim is the Cadence FastSPICE circuit simulator capable of predicting and validating timing, power and reliability of mixed-signal, complex digital and System-on-Chip (SoC) designs in advanced technology of  $0.13\mu m$  and below. It has a set of specialized reliability models (*AgeMos*) for HCI and NBTI simulation [15]. In the simulation, an *Age* parameter is calculated for each nMOSFET with the following formula:

$$Age(\tau) = \int_{t=0}^{t=\tau} \left[ \frac{I_{sub}}{I_{ds}} \right]^m \frac{I_{ds}}{W \cdot H} dt \quad (2.1)$$

where  $W$  refers to the channel width of the transistor,  $m$  and  $H$  are technology dependent parameters and determined from experiments,  $I_{sub}$  is the substrate current,  $I_{ds}$  is the drain-to-source current,  $\tau$  is the stress time. For pMOSFETs, the gate current  $I_{gate}$  instead of  $I_{sub}$  is used to determine the *Age* parameter. The degree of

device wearout has been experimentally found to be a function of this *Age* parameter for wide ranges of channel lengths and stress conditions, and the relationship has a plausible theoretical basis [16].

The simulation starts with device parameter extraction and modeling. From the SPICE model parameters of fresh devices, some other device parameters are added to accurately model  $I_{sub}$ . The next step is *AgeMos* extraction. Based on the *Age* parameter calculated from the fresh simulation, the *AgeMos* applies the degradation models, which can communicate with most SPICE-like simulators, to the aged circuit simulation. Reliability simulation with Virtuoso UltraSim is an iterative process, in which a large number of iterations are often needed in order to obtain accurate modeling results. The simulator can calculate and output the degradation results to predict the lifetime of each MOSFET within a circuit [17]. The overall simulation flow is illustrated in Fig.2.1.

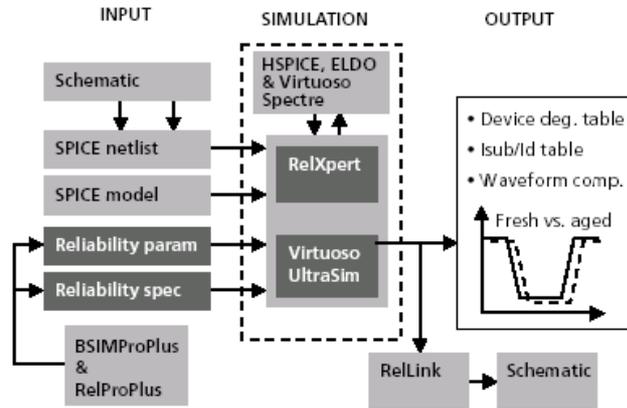


Figure 2.1: Hot carrier reliability simulation flowchart in Virtuoso UltraSim. Device wearout modeling is the focus of the reliability analysis [17].

The fundamental models and algorithms of reliability simulation realized in Virtuoso UltraSim found their origins in Berkeley Reliability Tool (BERT) which gives rise to many other reliability simulation tools. Most of these descendent tools are based on the same *Age* parameter modeling concept. The main advantages of these BERT-like tools are accuracy and SPICE compatibility, however, they also impose a burden on designers to correctly extract device’s fresh and degraded parameters and may lead to nonphysical trends, which prevents their popularity in reliability design process.

## 2.2.2 Reliability Simulation in Eldo

Eldo is a circuit simulator developed by Mentor Graphics which delivers all the capability and accuracy of SPICE-level simulation for complex analog circuits and SoC designs. In Eldo, the substrate current  $I_{sub}$  is not selected as the primary reliability parameter. In general, drain current  $I_d$ , threshold voltage  $V_t$  or transconductance  $g_m$  is often used as a degradation monitoring parameter, and the stress time resulting in 10% decrease of one of these monitoring parameters is arbitrarily set to the device lifetime. Degradation of  $I_d$  is a good monitor for digital circuits, while  $V_t$  shift is suitable for analog applications. Hot carrier reliability simulation in Eldo adopts  $I_d$  as degradation monitoring parameter and characterizes it with a compact  $\Delta I_d$  model, which directly models the difference of drain currents between fresh and aged devices.

There exist two competing mechanisms which lead to the obvious hot carrier

induced drain current variations between fresh and degraded devices: the deviation of  $I_d$  from its linear dependency of  $V_{ds}$  due to velocity saturation effects and the decreasing of  $\Delta I_d/I_d$  due to the reduction of charged interface states [18]. In Eldo, the  $\Delta I_d$  is modeled with equ.(2.2) to (2.5), which unify the subthreshold, linear and saturation regions with a simple relation for both forward and reverse operation modes [19]:

$$\frac{\Delta I_d}{I_d} = \frac{B_6(1 - e^{-B_1 V_{gs}}) + B_2}{1 + B_5(V_{gs} - B_3 V_t)} \times \frac{N_{it} L_{it}}{L_{eff}} \times \frac{1}{1 + \alpha(V_{ds} - V_{low}) + \beta V_{ds}} \quad (2.2)$$

$$V_{low} = A_3 V_{dsat} \quad (2.3)$$

$$\alpha = \frac{A_1}{1 + A_4(V_{gs} - V_t)^{A_2}} \quad (2.4)$$

$$\beta = A_5 V_{gs} + A_6 \quad (2.5)$$

where  $N_{it}$  is the interface trap density,  $L_{it}$  is the extension of the damage within the channel,  $L_{eff}$  is the effective channel length,  $V_{gs}$  is the gate-to-source voltage,  $V_t$  is the threshold voltage,  $V_{ds}$  is the drain-to-source voltage,  $V_{dsat}$  is the drain saturation voltage,  $A_1$  to  $A_6$  and  $B_1$  to  $B_6$  are model fitting parameters.

The same  $Age$  parameter defined by equ.(2.1) is incorporated to model the “age” of each transistor. The HCI aging process is simulated in an iterative way as depicted in Fig.2.2.

The period  $T_{age}$  at which the circuit performance is to be tested is divided into smaller time intervals  $T_1$ . The  $Age$  table is calculated at the end of each time interval and a new simulation with Eldo is carried forward. This process is repeated

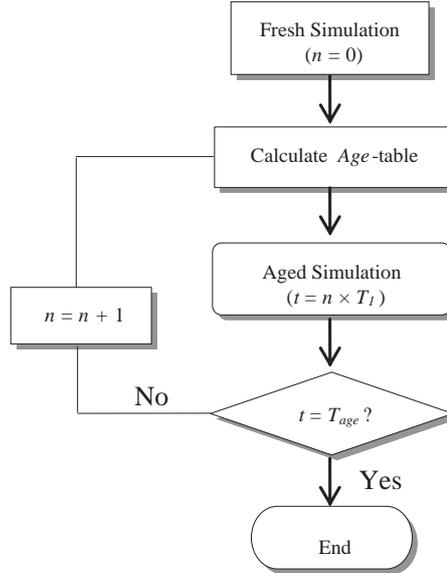


Figure 2.2: HCI reliability simulation in Eldo [19]. A large number of SPICE simulation iterations have to be carried out to obtain accuracy.

until  $T_{age}$  is reached. This iterative scheme can account for the gradual change of bias conditions as a result of device wearout.

The  $\Delta I_d$  modeling approach provides the possibility to have a relatively simpler parameter extraction process. It is suitable to model bi-directional stress and asymmetrical drain current behavior. However, because this approach also adopts both *Age* parameter and small-step iterative algorithm in the degradation simulation process, it inherits the same limitations of the BERT-like tools as discussed before.

## 2.3 Limitations and Improvements

Although the previous brief review reveals both the advantages and limitations of the contemporary reliability simulation tools, a further discussion is necessary for the sake of identifying the fundamental reasons for these limitations and understanding how MaCRO models and simulation algorithms overcome some of these limitations.

In reliability qualification practice, device lifetime or failure definition due to wearout mechanisms is quite arbitrary. A predefined shift in certain device parameter is often selected as the criterion for failure. Some examples are 10% reduction in  $I_{ds}$ , 10% decrease in  $g_m$ , or  $50mV$  shift in  $V_t$ . While these parameters' drift generally reflects device wearout degree, in real circuit applications, this treatment of device failures may not necessarily result in circuit failures. In order to establish a more realistic failure definition, Li et al at UIUC [20] proposed a new criterion which includes the estimation of both device local damage and circuit global degradation. Jiang et al at MIT [21] further used a 3% reduction in the critical path delay as the circuit-level failure criterion in the ripper-carry adder case study. Although significant improvement has been made in device failure modeling, no universally accepted method yet exists as what is device lifetime and how to assess impact of device failure on circuit-level reliability.

If device lifetime is defined as percentage or absolute drift in device parameters, then accurate calibration of the difference between fresh and degraded device parameters is indispensable for accurate circuit reliability simulation. However, pa-

parameter extraction for modeling individual device wearout to a satisfied accuracy is extremely tedious and difficult. In MaCRO, the focus of reliability analysis is circuit functionality rather than device wearout process, therefore, accurate characterization of each device parameter degradation is not necessary. A set of lifetime models for various wearout mechanisms are developed to identify those most degraded transistors in a circuit based on their terminal voltage and current waveforms. In this approach, normalize device lifetime values instead of absolute ones need to be predicted, as a result, device testing work and parameter extraction work are significantly alleviated.

Device wearout-focused reliability simulation tools treat various device wearout mechanisms with divide-and-conquer algorithm. Even though some of them like BERT have the capability to deal with EM, HCI and TDDB in the same environment [16, 22], each of these mechanisms is handled by a dedicated module with an assumption that every mechanism is independent from others. In reality, transistors in circuit expose to all kinds of stresses simultaneously and suffer from various wearout mechanisms which may interact with each other, as a result, the net effect of these combined mechanisms often leads to a precipitous degradation process. Another problem is some wearout processes are the synergic effects of two or more wearout mechanisms which have to be decoupled from each other in order to accurately characterize them individually. For example, both Channel Hot Carrier (CHC) injection and Biased Temperature Instability (BTI) mechanisms will contribute to interface trap generation which is the main reliability culprit in wearout process. Recently, some work has been done to uncover this interrelationship of

different wearout mechanisms. La Rosa et al at IBM [23] investigated the impact of both NBTI and CHC contributions to the device damage and proposed a methodology to decouple their effect. Yu et al at UCF [24] experimentally examined the interaction of NBTI with TDDB and HCI, and developed a transistor model to evaluate their combined effects on RF circuit performance degradation. Even with this progress, generally speaking, device wearout-focused reliability simulation methods cannot effectively deal with the combined effects of various mechanisms. In MaCRO, a set of circuit models are developed to characterize circuit failures due to multiple wearout mechanisms. This failure circuit modeling concept is not brand-new, some equivalent models have been explicitly proposed in literature (e.g. [24]). A thorough review of available circuit models for HCI, TDDB and NBTI is presented in the following chapters. Although these models are more or less rudimentary, to some extent they laid the foundation for further development of any advanced models. In MaCRO, the improved circuit models will be imported into SPICE netlists to substitute the most degraded transistors in the circuit. The SPICE simulation with these circuit models will reveal whether the circuit can survive from device wearout at any specific time.

Device wearout-focused reliability simulation tools only treat transistors suffering wearout mechanisms one by one in circuit. This is not accurate because neighboring devices also degrade at the same time and therefore influence terminal waveforms of the transistor under consideration. The effects of HCI on the operation of neighboring devices and circuits have been explored in [25]. For an nMOS transistor in a circuit, its threshold voltage will decrease and its subthreshold current

will increase due to excess substrate currents flowing in the neighboring MOS transistors and resulting from HCI and impact ionization effects. Some researchers have realized the problem of neglecting neighboring effects, but they turned to the other extreme case by taking into account all transistors' wearout effects at the same time. Obviously, these two cases are either inaccurate or inefficient. In a real circuit, different transistors operate at different biased points and therefore experience different stresses. Device lifetime is roughly exponentially dependent on these stress factors, which may lead to significant difference (sometimes even several orders of magnitude in difference, refer to Fig.1 in [26]) in device lifetime values. With MaCRO, by sorting normalized device lifetime values and only considering those transistors whose lifetimes are significantly smaller than others, designers may simultaneously obtain modeling accuracy and computational efficiency in addressing neighboring effects.

It is proved from IC reliability analysis that device DC lifetime is not sufficient to characterize circuit performance degradation. Therefore much work has been done to model device AC lifetime in circuits from static stress tests. Even though significant progress has been achieved in this field, due to the extreme complexity of device terminal waveforms in real circuits, there is still no convincing model available which is able to quantitatively predict device lifetime to a satisfied accuracy. Accurate and absolute value of device lifetime is theoretically important in reliability qualification, however, in engineering field, because of the statistical characteristics of device failure, an order of magnitude variation in predicted lifetime values is frequent and often tolerable. Compared with device wearout life, device

or circuit service life is extremely short, which makes the commonly adopted end-of-life characterization method rather ineffective in reliability analysis. End-of-life methods try to model the rising tail of bathtub curve, but more important and useful information is the level of failure rate in the middle part. An identifiable trend in reliability community is that hockey stick curve is gradually preempting bathtub curve in reliability analysis. With circuit complexity ever-increasing and the assumption that no wearout mechanism dominates in device, circuit failure distribution becomes more and more randomized. In this situation, circuit failure can be well approximated with an exponential distribution, and the failure rate ( $\lambda$ ) parameter solely characterizes the overall rate-of-failure process and reflects the level of reliability. This rate-of-failure concept is adopted in MaCRO to help develop lifetime models and predict circuit derating behaviors [27]. In developing lifetime models and determining add-on elements for circuit models, a quasi-static operation assumption is made which trades accuracy for simulation speed. This assumption conforms to the primary purpose of MaCRO: providing a simple tool for designers to make quick circuit performance and reliability evaluation. In literature, some advanced algorithms have been developed to address AC lifetime problem [28, 29, 30] which will be incorporated in MaCRO in future work.

In summary, the value of IC reliability simulation is not on determination of device and circuit absolute lifetime values, it should be able to provide chip designers simple guidelines to perform a quick circuit reliability evaluation, make appropriate tradeoffs between performance and reliability, and reduce product development cost and time. Reliability is unanimously regarded as a vital factor in successful prod-

uct development, however, reliability simulation has not been actively practiced in industry due to the reasons having been discussed. Most of the aforementioned limitations have been addressed in MaCRO, which treats circuit reliability from a different perspective by elevating reliability analysis from device wearout to circuit functionality. This circuit functionality-centered method integrates rate-of-failure concept, lifetime and circuit modeling techniques into a unified framework and provides designers an alternative in performing efficient circuit reliability simulation and analysis.

## 2.4 Assumptions and Justification

This section briefly explains the assumptions made in MaCRO. They are very important for people to understand the advantages of MaCRO models and simulation algorithms.

(1) Constant failure rate assumption. For the four wearout mechanisms being investigated (EM, HCI, TDDB and NBTI), even though they may not all follow exponential distribution, it has been justified that for a complex electronic system with multiple failure mechanisms, exponential distribution can be used to approximate overall failure rate. It is also proved that in the constant rate-based reliability analysis method, distribution of each failure mechanism is not absolutely necessary for predicting levels of system failure rate. Different trends and distributions of different mechanisms will be averaged out to a constant level of failure rate.

State-of-the-art VLSI devices are complex systems with millions of individual

transistors. Each transistor has at least a dozen of failure modes associated with it. Simulation shows that as the number of failure modes in a VLSI device increases to five or more, the Weibull shape parameter will shift toward unity unless all the modes have the same shape parameter and similar characteristic life. This simple observation implies that the failure rate of a VLSI approaches constant level as it becomes increasingly complex. With the further increase in complexity of a device, it will be difficult to distinguish any specific failure from others.

A good example of how increasing complexity results in a constant failure rate is the observation of the decrease in Weibull slope as the number of possible EM failure links in a device increases. EM is one of the most significant wearout failure mechanisms in electronic components. Each of those EM failure links has a strength associated with it which will vary with some distribution based on variables from design and process. The stress for each link is also a random variable. This series of random strengths, stresses, and the possibility of some lower strength links lead to a large spread of the probability distribution of the weakest link. With enough links the probability distribution function looks constant.

All these pieces of evidence prompt us to make the constant failure rate assumption. Constant failure rate-based reliability method for electronic components allows the VLSI manufacturers test parts under accelerated conditions assuming all failure mechanisms can be accelerated in approximately the same proportion. The resulting failure rate could then be extrapolated to operating conditions considering temperature, frequency and applied voltage. This extrapolation is the main consideration for product reliability engineers. Reliability design should be supported by

proper acceleration models, which can be verified through experimentally extracted values. Constant failure rate assumption not only leads to a simple system reliability model, but also inspires us to reevaluate the accelerated life test models currently used in industry.

(2) Equal contribution assumption. This means devices are properly designed with no dominant failure mechanism. As a result of improved knowledge of device failure mechanisms, electronic components are designed at the edge of “reasonable” life under tightly controlled specifications. Therefore, if any failure mechanism is more significant than others, specific design and manufacturing techniques will be developed to suppress this dominant failure. This assumption is the extension of the constant failure rate assumption. When no one failure mechanism dominates, all mechanisms are equally likely and the resulting failure distributions resemble constant rate processes.

(3) Linear superposition assumption. System failure rate is modeled as the sum of individual failure mechanisms. The result of this assumption is the Sum-Of-Failure-Rate (SOFR) system model. SOFR has been widely used in industry to model system and circuit reliability.

## 2.5 Summary of Lifetime and Circuit Models

The primary lifetime models and failure circuit models for each wearout mechanism are summarized in this section, detailed processes of developing these models are given in the following several chapters.

### 2.5.1 Hot Carrier Injection

The HCI lifetime model equation for nMOSFET is given by equ.(2.6):

$$t_f = A_{HCI} \left( \frac{I_{sub}}{W} \right)^{-n} \exp\left( \frac{E_{aHCI}}{\kappa T} \right) \quad (2.6)$$

where  $I_{sub}$  is substrate leakage current,  $E_{aHCI}$  is the activation energy,  $W$  is the channel width,  $\kappa$  is Boltzmann's constant,  $T$  is temperature,  $n$  is a process related constant,  $A_{HCI}$  is the model prefactor. For pMOSFET HCI lifetime model, the gate leakage current  $I_{gate}$  replaces  $I_{sub}$  in equ.(2.6).

The HCI circuit model for nMOSFET is illustrated in Fig.2.3, which is based on the  $\Delta R_d$  model [31] with some improvements. The inclusion of  $\Delta R_d$  emulates the degradation of drain-to-source current  $I_{ds}$ . Both interface trap generation and oxide charge trapping contribute to the increase in  $\Delta R_d$  value. The contribution of oxide charge trapping to device wearout is neglected in the original  $\Delta R_d$  model [31], but recent experimental work and the SRAM simulation results presented in Chapter 8 prove that oxide trapped charge is also a major contributor to device wearout.

### 2.5.2 Time Dependent Dielectric Breakdown

The TDDB lifetime model equation for nMOSFET is based on the work by Wu et al at IBM [32, 33, 34, 35] and given by equ.(2.7):

$$t_f = A_{TDDB} \left( \frac{1}{A} \right)^{\frac{1}{\beta}} F^{\frac{1}{\beta}} V_{gs}^{a+bT} \exp\left( \frac{c}{T} + \frac{d}{T^2} \right) \quad (2.7)$$

where  $A = W \times L$  is the device gate oxide area,  $\beta$  is Weibull slope parameter,  $F$  is cumulative failure percentile at use condition,  $V_{gs}$  is gate-to-source voltage,  $T$  is

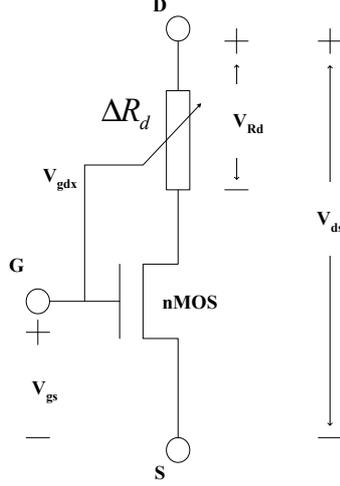


Figure 2.3: HCI circuit model in MaCRO. In the model:  $V_{gdx} = V_{gs} - V_t - V_{ds}$  and  $V_{Rd} = I_{ds}\Delta R_d$ .  $V_t$  is threshold voltage and  $I_{ds}$  is the current from node  $D$  to  $S$ .

temperature,  $a$ ,  $b$ ,  $c$ , and  $d$  are model fitting parameters determined from experimental work,  $A_{TDDb}$  is the model prefactor. Note that  $a + bT$  is always negative. Equ.(2.7) is the result of various TDDb experimental observations including power law voltage acceleration, non-Arrhenius temperature acceleration, weakest-link area scaling law and so on.

The TDDb circuit model for nMOSFET is illustrated in Fig.2.4, in which two split transistors imitate the channel separation by oxide breakdown path, and the voltage-dependent current source  $I_{OX}$  physically represents the conduction mechanism of hard breakdown path across the oxide. Fig.2.4 is the model for gate-to-channel breakdown scenario, which is a much more frequent statistical event than gate-to-diffusion breakdowns. Gate-to-diffusion breakdowns have more severe damages on device operation. A simple gate-to-source or gate-to-drain parasitic resistance is used for modeling gate-to-diffusion breakdown effects.

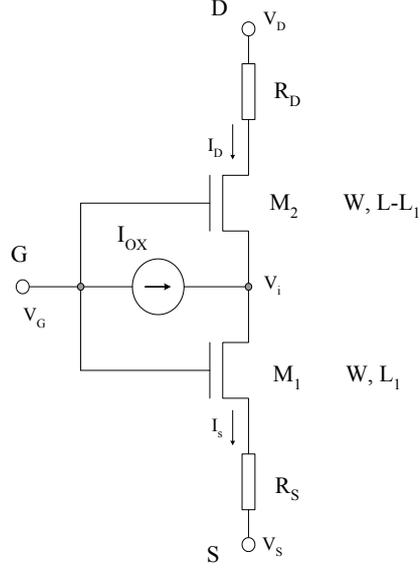


Figure 2.4: TDDB circuit model in MaCRO.  $I_{OX} = I_S - I_D$  is a voltage-dependent current source representing breakdown path current injection effect.  $R_D$  and  $R_S$  characterize the resistance in the source and the drain extensions, respectively.  $L_1$  represents breakdown location in terms of the source edge.

### 2.5.3 Negative Bias Temperature Instability

The NBTI lifetime model equation for pMOSFET is based on the physics and statistics model proposed by Zafar et al at IBM [36, 37] and shown as equ.(2.8):

$$t_f = A_{NBTI} V_{gs}^{-\frac{1}{\beta}} \left[ \frac{1}{1 + 2 \exp(-\frac{E_1}{\kappa T})} + \frac{1}{1 + 2 \exp(-\frac{E_2}{\kappa T})} \right]^{-\frac{1}{\beta}} \quad (2.8)$$

where  $\beta$  is model fitting parameter,  $E_1$  is a material related constant,  $E_2$  is a material and oxide field dependent parameter,  $V_{gs}$  is gate-to-source voltage,  $A_{NBTI}$  is the model prefactor.

The NBTI circuit model for pMOSFET is illustrated in Fig.2.5, in which NBTI-induced pMOSFET threshold voltage increase is modeled as absolute gate-to-

source voltage decrease. Gate tunneling current flowing through the gate resistance  $R_G$  leads to the increase of voltage at point  $G'$ . This corresponds to the decrease of pMOSFET absolute gate-to-source voltage and therefore mimics the threshold voltage degradation effect. Gate tunneling current is modeled with two voltage controlled current sources following a simple formula:  $I = KV^P$ , which is actually a power-law leakage current model for TDDDB effect. The exponent  $p$  varies from 5 to 2 as the degradation level increases, and  $K$  reflects the “size” of the oxide breakdown spot.

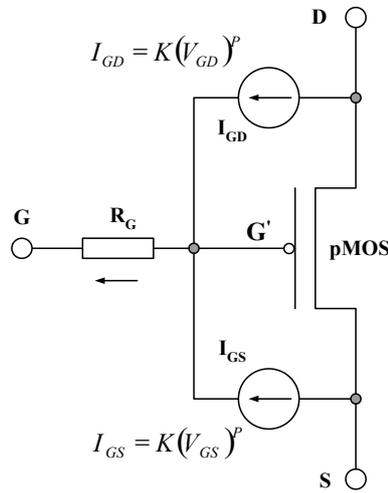


Figure 2.5: NBTI circuit model in MaCRO. The inclusion of  $I_{GD}$  and  $I_{GS}$  inherently accounts for oxide breakdown effects and also supplies leakage currents for  $R_G$  whose voltage drop is equivalent to pMOSFET threshold voltage degradation.

## 2.6 Reliability Prediction and Simulation Algorithms

The MaCRO lifetime and circuit models can be tailored for different purposes of reliability analyses: if the circuit lifetime is of primary interest, designers can manipulate the lifetime models to accurately predict device and circuit lifetimes after properly extracting all model parameters; if the circuit functionality is of primary interest, they can quickly identify weakest devices with normalized lifetime calculation and incorporate the circuit models to simulate circuit operations and check functionality at any interested time. The first kind of above analyses, i.e. lifetime prediction, can be further diversified into three different categories: when all lifetime model parameters are obtained from experimental work, MaCRO can estimate circuit lifetime and failure rate; if process parameters of future technologies are projected, MaCRO can predict reliability trends over generations and identify potential reliability showstopper; if all derating factor model parameters are calibrated, MaCRO can be used in voltage and temperature derating analysis aiming at reliability enhancement. Except for derating modeling which is addressed in Chapter 7, the MaCRO flowcharts and simulation algorithms for other reliability analysis methods are presented in this section.

### 2.6.1 Circuit Lifetime and Failure Rate Prediction

The lifetimes of each transistor in a circuit with respect to different wearout mechanisms have been given by equ.(2.6), (2.7), and (2.8). To obtain the lifetime of the entire circuit, one need to combine the effects of different mechanisms across dif-

ferent structures. This requires information of time-dependent lifetime distribution for each mechanism. In engineering applications, the FIT value is normally used to quantify product reliability, which represents the number of failures per  $10^9$  device-hours of stress testing. Most FIT calculation methods only apply to systems with constant failure rate for each failure mechanism, so special treatment is required for other systems having failures with time-variant characteristics [38].

With further development of deep submicron technologies, integrated circuits become increasingly complex, and both the physical dimensions and logic functions of each unit are being exploited to their limits. Every unit is prone to fail in a shorter time, and if it does fail, the system will be greatly impaired or even fails at the same time. Therefore a complex integrated circuit can be approximated with a competing failure system, i.e. a series failure system. The main feature of a series failure system is that the first failure of individual unit will lead to the failure of the whole system, therefore, system reliability function is the multiplication of individual reliability functions. Another practical approximation is that each failure mechanism could be treated with exponential distribution. In this way, the failure rate of each failure mechanism is approximated as a constant. With these two assumptions, one can apply the standard Sum-Of-Failure-Rates (SOFR) model to system failure rate calculation from its individual failure mechanisms [39].

According to SOFR model, the Mean Time To Failure (MTTF) of a circuit composed of  $n$  units can be related to the lifetime of each unit ( $MTTF_{ij}$ ) due to

each of the  $m$  individual failure mechanisms with equ.(2.9):

$$\frac{1}{MTTF} = \sum_{i=1}^m \sum_{j=1}^n \frac{1}{MTTF_{ij}} \quad (2.9)$$

The FIT is interchangeable with MTTF according to its definition for a constant failure rate system:

$$FIT = \frac{10^9}{MTTF} \quad (2.10)$$

If all the parameters of the lifetime models presented in the above section have been extrapolated from device testing work, from equ.(2.6) to (2.10), the MTTF and FIT of the circuit can be obtained.

The MaCRO flowchart of the circuit lifetime and failure rate prediction process is depicted in Fig.2.6. With minor modifications, Fig.2.6 is also applicable to reliability trends analysis for future technologies. For example, if device SPICE model parameters and MaCRO model fitting parameters of future technologies are reasonably projected, designers can predict the reliability trend of 65nm process and beyond in light of the wearout mechanisms being discussed. This kind of reliability analysis is very important for any further CMOS scaling.

A more detailed version of MaCRO flow for lifetime and failure rate prediction is attached in Appendix B.

A natural derivative of the above circuit lifetime and failure rate prediction algorithm is on derating stress factors, i.e. voltage and temperature, for higher reliability. The derating methods with MaCRO lifetime models is discussed in Chapter 7 where EM effect is also included after EM lifetime model being given in Chapter 6.

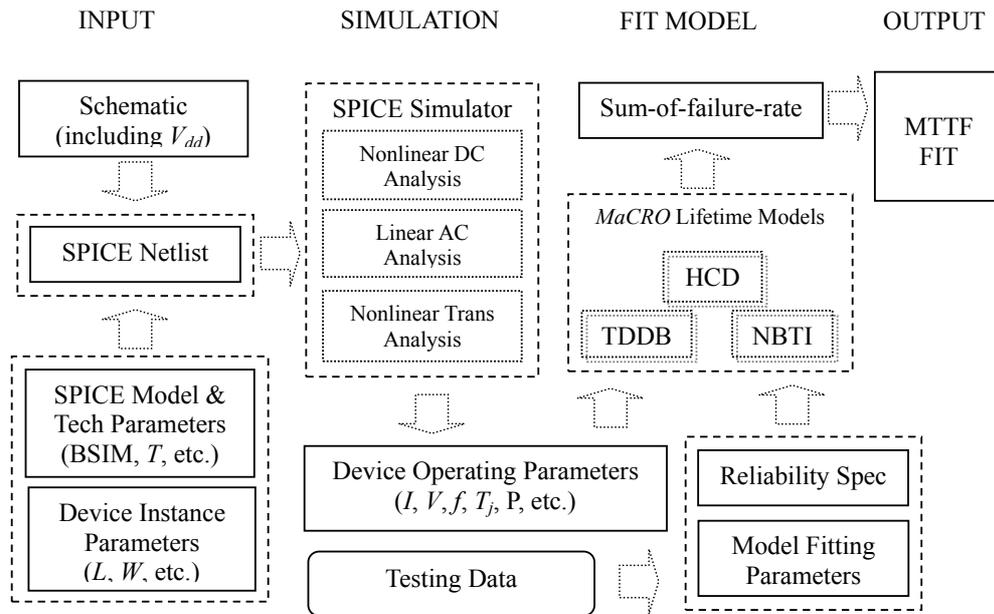


Figure 2.6: Flowchart of device and circuit lifetime and failure rate prediction process with MaCRO lifetime models. SPICE simulation predicts device terminal voltage and current stress profiles, and model fitting parameters are determined from device testing work.

The last point deserving special attention in lifetime prediction is the accuracy problem limited by quasi-static assumption which neglects HCI and TDDB AC acceleration effects and NBTI dynamic recovery effects. In estimating device terminal voltage and current stress profiles with SPICE, even though device operation is dynamic, for simplicity only time average values of these terminal waveforms are calculated. If terminal waveforms are clean and regular, duty cycle instead of time averaging method can be applied to improve accuracy. The waveform averaging method based on duty cycles is used in the SRAM reliability simulation work which is presented in Chapter 8. In general, there is no accurate model for dynamic

stress analysis, this, plus the complexity in extracting all model parameters, limits the applicability of MaCRO in lifetime prediction. In order to overcome this limitation, MaCRO shifts focus of reliability analysis from absolute lifetime prediction and device wearout to normalized lifetime calculation and circuit functionality.

## 2.6.2 Circuit Reliability Simulation Algorithm

MaCRO circuit reliability simulation algorithm is fundamentally a two-step SPICE simulation process. First, SPICE simulation is performed without considering any wearout mechanisms. From the first simulation run, terminal voltage and current stress profiles for each transistor can be obtained. Then, the lifetime models for HCI, TDDB and NBTI are called to compute every device's normalized lifetime for each mechanism, and a set of device tables, ranked by normalized lifetime values, are generated for designers to identify the most degraded transistors.

After identifying the most degraded transistors, MaCRO calls SPICE engine again. The second round SPICE simulation is performed by substituting those identified transistors with corresponding circuit models individually or jointly depending on whether a specific transistor experiences single or multiple wearout mechanisms. The model parameters for each circuit are calculated with a dedicated Matlab routine which contains both predefined device and process parameters as well as user-input parameters. These Matlab programs are listed in Appendix A.

From the second SPICE simulation run, circuit performance and functionality are expected to change due to the incorporation of the circuit models which may have

changed circuit internal connections, biasing networks and local topology. The circuit functionality may or may not be preserved depending on the magnitude of these additional circuit elements. After very limited times of SPICE simulation, circuit functional lifetime and failure behaviors can be easily predicted and characterized. With this information, circuit designers can quickly perform design iterations to improve circuit reliability if circuit functional lifetime falls short of specifications. They can also work on specific devices in circuit, sweep their circuit model parameters, and find the critical values corresponding to specific device wearout level at which circuit function fails. From this kind of analysis, designers can explore circuit reliability margins, and make appropriate performance and reliability tradeoffs. The pseudocode of the above process is illustrated in Fig.2.7.

### **MaCRO Circuit Reliability Simulation Algorithm**

**Inputs:** Model fitting parameters;  
SPICE model parameters;  
Circuit schematic/netlist;

**Start:** Fresh SPICE simulation;  
Calculate average values of device's operating parameters;

**Weakest Devices Identification:**  
For HCI:  
    Call HCI accelerated lifetime model;  
    Sort normalized lifetimes and identify weakest devices;  
For TDDB:  
    Call TDDB accelerated lifetime model;  
    Sort normalized lifetimes and identify weakest devices;  
For NBTI:  
    Call NBTI accelerated lifetime model;  
    Sort normalized lifetimes and identify weakest devices;

**Simulation with Failure Models:**  
HCI effects: (1) Initialize:  $Tr = 0$ ,  $Tw = \text{large value}$ ,  $t = (Tw - Tr)/2$ ;  
(2) Calculate HCI circuit model parameters at time  $t$ ;  
(3) Insert HCI circuit models in netlist and perform SPICE simulation;  
(4) if  $(Tw - Tr) < \Delta$ , go to (5);      $\Delta$  is a predefined small value.  
    else check circuit functionality:  
        if correct, set  $Tr = t$ , and  $t = Tr + (Tw - Tr)/2$ , then repeat (2) to (4);  
        if failed, set  $Tw = t$ , and  $t = Tr + (Tw - Tr)/2$ , then repeat (2) to (4);  
(5) Set  $Ta = t$ ;      $Ta$  is circuit HCI lifetime.  
TDDB+HCI: (6) Initialize:  $Tr = 0$ ,  $Tw = Ta$ ,  $t = (Tw - Tr)/2$ ;  
(7) Calculate TDDB+HCI circuit model parameters at time  $t$ ;  
(8) Insert TDDB+HCI models in netlist and perform SPICE simulation;  
(9) Repeat step (4) with inclusion of TDDB models;  
(10) Set  $Tb = t$ ;      $Tb$  is circuit TDDB+HCI lifetime.  
NBTI+TDDB+HCI: (11) Initialize:  $Tr = 0$ ,  $Tw = Tb$ ,  $t = (Tw - Tr)/2$ ;  
(12) Calculate NBTI+TDDB+HCI model parameters at time  $t$ ;  
(13) Perform SPICE simulation with NBTI+TDDB+HCI models;  
(14) Repeat step (4) with inclusion of NBTI+TDDB models;  
(15) Set  $Tc = t$ ;      $Tc$  is circuit NBTI+TDDB+HCI lifetime.  
Result: circuit functions until  $Tc$ , and fails at times beyond it.

**Circuit Reliability Analysis:**  
SPICE DC analysis at time  $Tc$ ;  
SPICE AC analysis at time  $Tc$ ;  
SPICE XF analysis at time  $Tc$ ;  
...

Figure 2.7: MaCRO circuit reliability simulation algorithm.

## Chapter 3

### Hot Carrier Injection Effect and Models

#### 3.1 Introduction

Hot Carrier Injection (HCI) is the phenomenon that carriers at MOSFET's drain end gain sufficient energy to inject into the gate oxide and cause degradation of some device parameters. Channel carriers become "hot" as they shoot out from the source of a MOSFET, accelerate in the channel, and experience impact ionization near the drain junction due to high lateral electric field [40]. Under favorable conditions, some high energy electrons and/or holes produced by the impact ionization are re-directed and accelerated toward the interface of oxide and silicon surface. A few "lucky" carriers overcome the surface energy barrier, inject into the oxide, and generate interface states and oxide charges, which are the main mechanisms for degradation of some MOSFET parameters such as channel mobility, threshold voltage, transconductance and drain saturation current. The shifts in threshold voltage and transconductance are proportional to the average trap density, which in turn is inversely proportional to the effective channel length [41]. Therefore, reducing the channel length will exacerbate hot carrier effect. For future CMOS technologies, even the power supply voltage will be reduced to 1V or below, HCI is still

a significant reliability concern due to continuous scaling of device channel lengths [42].

Among the three wearout mechanisms considered in MaCRO, HCI is the most thoroughly investigated one, and quite a few hot carrier lifetime models and SPICE failure macro models have been proposed in the past two decades. While some of the HCI lifetime models are based on the simple drain voltage accelerating law, most other successful lifetime models characterize HCI effect with peak substrate current for nMOSFETs and peak gate current for pMOSFETs. These semi-empirical models are valid at least down to  $0.25\mu m$  technology. In the generations beyond ( $0.25\mu m \sim 0.07\mu m$ ), research has shown that existing lifetime models remain more or less applicable at low voltages [43]. In order to characterize HCI effects in circuit environment, many HCI SPICE macro models have been proposed and integrated into reliability simulation tools. Some of these SPICE macro models are reviewed in this chapter before introducing the improved  $\Delta R_d$  HCI circuit model adopted in MaCRO.

### 3.2 Accelerated Lifetime Model

Most HCI lifetime models are based on the “lucky electron” model, upon which the hot carrier stress on an nMOSFET, in terms of generated interface traps  $\Delta N_{it}$ , can be related to the electric field  $E_m$  at the drain, the drain-to-source current  $I_{ds}$  and stress time  $t$  in a simple power-law relation [44]:

$$\Delta N_{it} = C_1 \left[ \frac{I_{ds}}{W} \exp\left(-\frac{\Phi_{it,e}}{q\lambda_e E_m}\right) t \right]^n \quad (3.1)$$

where  $W$  is the channel width,  $\Phi_{it,e}$  is the critical energy for electrons to create an interface trap ( $\Phi_{it,e} = 3.7eV$  [45]),  $\lambda_e$  is the hot-electron mean-free path ( $\lambda_e = 6.7nm$  [46]),  $C_1$  is a process constant. The dynamics of interface trap generation is similar to the rate of thermal oxide growth: at initial stage, interface trap generation rate is reaction limited, therefore,  $\Delta N_{it} \propto t$  and  $n = 1$ ; at later stage, the generation is diffusion limited, then  $\Delta N_{it} \propto t^{1/2}$  and  $n = 0.5$ . The overall process is the compromised result of these two competing processes and as a result the parameter  $n$  falls within the range between 0.5 and 1 [45]. In MaCRO, the typical value of  $n$  is set to 0.65.

The most important parameter in equ.(3.1) is the electric field  $E_m$  which cannot be determined accurately by simple calculation. A semi-quantitative analytical  $E_m$  model has been given in [45]:

$$E_m = \frac{V_{ds} - V_{dsat}}{\sqrt{3t_{ox}x_j}} \quad (3.2)$$

where  $t_{ox}$  is the gate oxide thickness,  $x_j$  is the drain junction depth.  $\sqrt{3t_{ox}x_j}$  is the characteristic length which models the effective thickness of the channel “pinchoff” region whose typical values are within  $\sqrt{100nm}$  to  $\sqrt{300nm}$ . The factor 3 in  $\sqrt{3t_{ox}x_j}$  derives from the ratio of  $\epsilon_{s_i}/\epsilon_{s_iO_2}$  [47]. In MaCRO, the default value of  $\sqrt{3t_{ox}x_j}$  is  $10nm$ .

In equ.(3.2),  $V_{dsat}$  is the potential at the channel “pinchoff” point. There are many models for  $V_{dsat}$ , among which the simplest one is  $V_{dsat} = V_{gs} - V_t$ , where  $V_{gs}$  is gate-to-source voltage and  $V_t$  is the threshold voltage. For short channel devices,

$V_{dsat}$  is channel length ( $L$ ) dependent, and the relation is often modeled as [45]:

$$V_{dsat} = \frac{(V_{gs} - V_t)LE_{cr}}{V_{gs} - V_t + LE_{cr}} \quad (3.3)$$

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

In the above discussion, the only unknown parameter in equ.(3.1) is the coefficient  $C_1$  which is a process determined constant. For each technology, it only needs to be characterized once. The typical values of  $C_1$  are within  $1.9 \sim 2$  according to [46] (on pp.67).

Besides the interface trap generation model given by equ.(3.1), the other two important models for hot carrier effects are substrate current ( $I_{sub}$ ) model and gate current ( $I_{gate}$ ) model:

$$I_{sub} = C_2 I_{ds} \exp\left(-\frac{\Phi_i}{q\lambda_e E_m}\right) \quad (3.4)$$

and

$$I_{gate} = C_3 I_{ds} \exp\left(-\frac{\Phi_b}{q\lambda_e E_m}\right) \quad (3.5)$$

where  $\Phi_i$  is the minimum energy (in electronvolt) for a hot electron to create an impact ionization ( $\Phi_i = 1.3eV$ ),  $\Phi_b$  is the barrier energy (also in electronvolt) at the  $Si-SiO_2$  interface. The formula for  $\Phi_b$  is given by equ. (3.9) in [46] (on pp.61). The constants  $C_2$  and  $C_3$  are given in [45] as  $C_2 = 2$  and  $C_3 = 2 \times 10^{-3}$ .

By defining the device hot carrier lifetime  $t_f$  as the time to reach a fixed amount of interface trap density, we can combine equ.(3.1) and equ.(3.4) into a very

useful lifetime equation:

$$\frac{t_f I_{ds}}{W} = C_4 \left[ \frac{I_{sub}}{I_{ds}} \right]^{-\Phi_{it,e}/\Phi_i} \quad (3.6)$$

Equ.(3.6) is used in many hot carrier reliability simulation tools derived from BERT [48]. From this equation, a very simple lifetime model for HCI can be obtained:

$$t_f = C_5 \exp\left(\frac{\theta}{V_{ds}}\right) \quad (3.7)$$

where  $C_5$  and  $\theta$  are technology related constants whose values are determined from accelerated tests,  $V_{ds}$  is the drain-to-source voltage. The power of equ.(3.7) is that it relates a device's HCI lifetime to only one operating parameter which can be directly calibrated from SPICE simulation. The main problem for this simple relation is that it is only valid for a small range of gate voltages near the maximum substrate current [44], which corresponds to the stress conditions that gate voltage is close to the middle value of drain voltage.

In order to take into account realistic hot carrier stress profiles in circuit environment, a more general lifetime model is incorporated in MaCRO which relies on the substrate current model.  $I_{sub}$  has been identified as the best hot carrier reliability monitor for nMOSFETs. According to [49], the device parameter degradation due to HCI can be modeled as:

$$\Delta P = C_6 \left( \frac{I_{sub}}{W} \right)^{\alpha} t^{\beta} \quad (3.8)$$

where  $I_{sub}/W$  is the normalized substrate current,  $\alpha$ ,  $\beta$  and  $C_6$  are technology related constants.

Temperature acceleration is often treated as a minor effect in most HCI models, however, in order to consider possible large temperature excursions, MaCRO includes temperature acceleration effect based on the HCI lifetime model given in [40]. The combination of temperature effect and equ.(3.8) produces a more comprehensive HCI lifetime model:

$$t_f = A_{HCI} \left( \frac{I_{sub}}{W} \right)^{-n} \exp\left( \frac{E_{aHCI}}{\kappa T} \right) \quad (3.9)$$

where  $E_{aHCI}$  is the apparent activation energy (the typical value of  $E_{aHCI}$  is within  $-0.1eV$  to  $-0.2eV$ , but it can be negative or positive depending on device technology),  $W$  is the device gate width,  $\kappa$  is Boltzmann's constant ( $\kappa = 8.62 \times 10^{-5} eV/K$ ),  $T$  is temperature in Kelvin,  $n$  is a technology dependent constant,  $A_{HCI}$  is the model prefactor. In MaCRO, the default values for  $n$  and  $E_{aHCI}$  are  $n = 1.5$  and  $E_{aHCI} = -0.15eV$ , respectively.

There are two ways to determine  $I_{sub}$ : one is from equ.(3.4), the other way is from BSIM3 model equations as follows:

$$I_{sub} = \frac{\alpha_0 + \alpha_1 L_{eff}}{L_{eff}} V_{ds}' \exp\left( \frac{-\beta_0}{V_{ds}'} \right) \frac{I_{ds0}(1 + V_{ds}'/V_A)}{1 + R_{ds} I_{ds0}/I_{dseff}} \quad (3.10)$$

$$V_{ds}' = V_{ds} - V_{dseff} \quad (3.11)$$

The meaning of the above model parameters is given in BSIM3 Model User Manual [50]. This BSIM3  $I_{sub}$  model is quite similar to the  $I_{sub}$  model proposed in iProbe-d [51], therefore, the iProbe-d  $I_{sub}$  model is an alternative if some SPICE simulator does not support BSIM3  $I_{sub}$  calculation.

The degradation of pMOSFETs under hot carrier stress is becoming one of the important contributors to circuit reliability. The hot carrier induced pMOSFET degradation effects on circuit performance is different from those of nMOSFET in that they may lead to reverse shifts of pMOSFET parameters (in terms of directions of parameter shifts in nMOSFET) due to significant negative charge trapping in oxide rather than excessive interface trap generation. The circuit performance degradation can be characterized more accurately if pMOSFET HCI effect is also considered. Even though the wearout dynamics and device parameter degradation trends of pMOSFETs are different from those of nMOSFETs, with minor modifications, the above nMOSFET's lifetime models can be applied to pMOSFETs and given as.

$$t_f = A_{HCI,p} \left( \frac{I_{gate}}{W} \right)^{-m} \exp\left( \frac{E_{aHCI,p}}{\kappa T} \right) \quad (3.12)$$

where  $E_{aHCI,p}$  is the apparent activation energy ( $E_{aHCI,p}$  is within  $-0.1eV \sim -0.2eV$ ),  $W$  is the device gate width.  $m$  and  $A_{HCI,p}$  are technology related constants, whose default values in MaCRO are  $m = 12.5$  and  $E_{aHCI,p} = -0.15eV$ , respectively. The  $I_{gate}$  is given by equ.(3.5).

In developing MaCRO HCI lifetime models, a quasi-static assumption is made which averages device dynamic operation parameters (e.g.  $I_{ds}$ ,  $V_{ds}$ ,  $V_{gs}$ ) in terms of simulation time, therefore,  $I_{sub}$  and  $I_{gate}$  in equ.(3.9) and (3.12) are average values calculated from equ.(3.4) and equ.(3.5), respectively. The same assumption also applies to TDDB and NBTI lifetime models. The method to improve accuracy in estimating ‘‘average values’’ of these operation parameters is briefly discussed in the

previous chapter.

### 3.3 Failure Equivalent Circuit Model

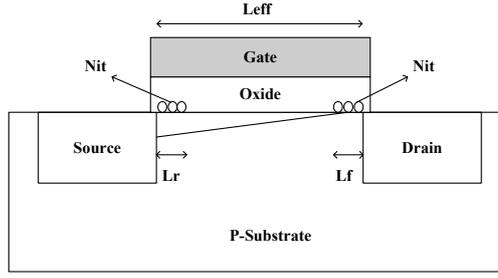
In order to account for the effect of device hot carrier damage on circuit functionality and reliability, the device-level lifetime models have to be extended to circuit-level applications. The bridge connecting the gap between device wearout degree and circuit performance drift is no doubt the circuit models. The underlying concept of the circuit models is modeling degradation of device parameters with some additional lumped circuit elements (resistors, transistors or dependent current sources, etc.) to capture the behavior of a damaged MOSFET in circuit operation environment. The values of these additional lumped elements are determined by device wearout parameters (such as  $\Delta N_{it}$ ) which are time dependent and by device terminal voltage and current waveforms, therefore, at any time  $t$ , values of these lumped elements can be predicted accurately and their magnitude reflects the device wearout degree. The larger the magnitude of these values, the severer the damage to circuit functionality. As a result, circuit designers can quickly analyze circuit reliability behaviors at any given time with these circuit models.

Several HCI circuit models have been developed in the past years and some of them have been built into commercial reliability simulation tools. Almost all circuit models are based on SPICE simulation platform which is a de facto tool in circuit design. In this section, some of these circuit models is briefly reviewed, followed by the introduction of the HCI circuit model implemented in MaCRO.

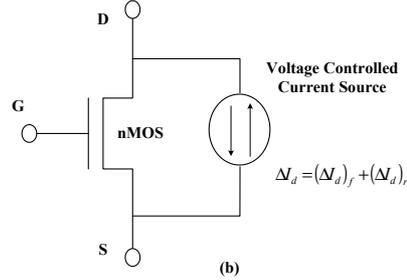
BERT is up to now the most successful circuit reliability simulation tool. BERT directly models nMOSFET hot carrier damage in drain current degradation. The drain current degradation,  $\Delta I_d$ , results from channel mobility degradation, which again results from HCI-induced interface traps  $\Delta N_{it}$ .  $\Delta N_{it}$  is modeled in terms of the famous *Age* parameter introduced in the previous chapter. In BERT,  $\Delta I_d$  is implemented as an asymmetrical voltage controlled current source in parallel with the original nMOSFET. The pMOSFET HCI effect is modeled with the concept of channel shortening and drain resistance increase [48]. The BERT  $\Delta I_d$  model is shown in Fig.3.1, which captures the asymmetrical forward and reverse  $I - V$  characteristics and allows simulation of devices undergoing bi-directional stresses (such as devices in a transmission gate).

The detailed  $\Delta I_d$  model equations and parameters are defined in [53]. The main contribution of BERT  $\Delta I_d$  model is the ability to characterize bi-directional hot carrier stress effects, however it requires extraction of six process parameters from device testing, which is a non-trivial work.

Experiments have proved that HCI-induced interface traps in nMOSFET is localized above the channel near the drain junction. More specifically, these interface traps are localized in the vicinity within  $100nm$  from the drain [45]. Based on this observation, Leblebici et al at UIUC [46, 54] developed a two-transistor HCI circuit model which consists of an HCI damaged parasitic transistor with fixed channel length  $L_2$  ( $L_2 \approx 0.1\mu m$ ) in series connection with the original transistor whose channel length is shrunk to  $L - L_2$ . The primary assumption for this model is that all generated interface traps are occupied with electrons, which equals to considering



(a)



(b)

Figure 3.1: BERT nMOSFET HCI circuit model. (a) Bidirectional interface trap generation near both drain and source.  $L_f$  and  $L_r$  represent forward and reverse hot carrier damaged regions. (b) HCI drain current  $\Delta I_d$  circuit model [52].

only negative fixed charge. The model is illustrated in Fig.3.2.

From Fig.3.2 (a), the interface trapped charge  $Q_{it}$  due to HCI can be readily derived as:

when  $(0 \leq x < L_1)$

$$Q_{it}(x) = 0 \quad (3.13)$$

when  $(L_1 \leq x < L)$

$$Q_{it}(x) = \frac{Q_M}{L_2}(x - L_1) \quad (3.14)$$

where  $Q_M$  denotes the largest interface charge,  $L_1 = L - L_2$ , and  $L_2$  represents the

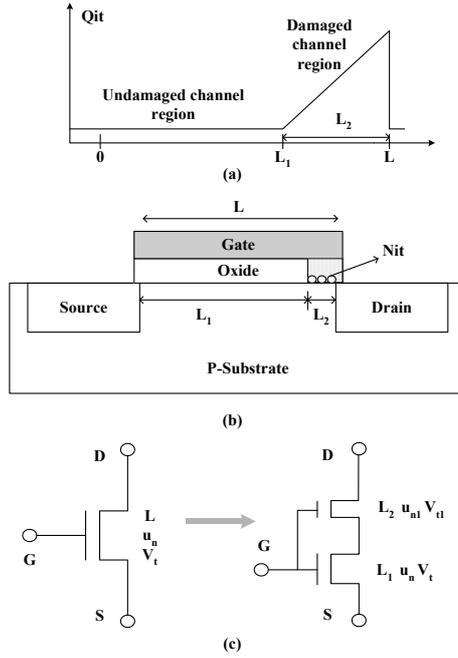


Figure 3.2: UIUC nMOSFET HCI two-transistor series model. (a) Triangular oxide charge distribution profile used in model derivation. (b) Cross-sectional view of nMOSFET with hot carrier damage,  $L_2$  is damaged channel region. (c) Two-transistor series circuit model. The parasitic transistor has different channel mobility and threshold voltage with the channel length  $L_2$  set to  $0.1\mu m$  [46, 51, 55].

length of the damaged channel region. This two-transistor model characterizes the amount of hot carrier damage with only two parameters  $Q_M$  and  $L_2$ , therefore, the model parameter extraction work is greatly reduced. The drawbacks of this model are in two aspects: the triangular charge density distribution is over simplified, and it is not easy to extrapolate  $Q_M$  value.

Up to now, the simplest HCI circuit model is the Hot Carrier Induced Series

Resistance Enhancement Model (HISREM), also named  $\Delta R_d$  model, which is proposed by Hwang et al at Oregon State University [56]. Based on the fact that the increase of HCI-induced series drain resistance is due to the injection of hot carriers close to the drain edge, a series resistance  $\Delta R_d$  added to the drain of the nMOSFET can reflect the process of hot carrier induced interface trap generation and therefore accounts for the channel mobility reduction and threshold voltage drifts. HISREM consists of a voltage dependent drain resistor  $\Delta R_d$  connected in series with the original nMOSFET.  $\Delta R_d$  is a function of the applied voltages and the hot carrier induced interface trapped charge  $\Delta N_{it}$ . The behavior of the damaged nMOSFET is emulated by the original undamaged device operated with a reduced drain-to-source voltage which is controlled by this additional drain resistor  $\Delta R_d$ . Because  $\Delta N_{it}$  is a time dependent parameter,  $\Delta R_d$  model is able to predict drain current degradation at any given time. HISREM is also capable of modeling self-limiting effects of hot carrier damage because the increase in series drain resistance of an nMOSFET suppresses hot carrier stress. The most advantageous feature of HISREM model is that only one parameter,  $\Delta N_{it}$ , needs to be extrapolated from device testing work. Consequently, HISREM model can be easily used by circuit designers to perform an expeditious reliability analysis.

HCI circuit model in MaCRO is based on the above  $\Delta R_d$  model with some improvements. The major improvement is that  $\Delta R_d$  value is considered to be determined by both interface trapped charge  $\Delta N_{it}$  and oxide trapped charge  $\Delta N_{ox}$ . The contribution of  $\Delta N_{ox}$  to device wearout is often neglected, but recent experimental work recognizes that they can account for some of the observed enhanced

degradation effects in nMOSFETs which could not be explained solely by  $\Delta N_{it}$  generation.

MaCRO HCI circuit model is illustrated in Fig.3.3. The derivation of  $\Delta R_d$  is carried out under the assumptions that (1) all interface traps are acceptor-like and occupied by electrons, and (2) channel mobility degradation,  $\mu$ , is caused by both  $\Delta N_{it}$  and  $\Delta N_{ox}$ . The assumption (1) means the net charge in interface traps is a fixed negative charge for nMOSFET in strong inversion operation. The assumption (2) leads to the following equation:

$$\mu = \frac{\mu_0}{1 + \alpha \Delta N} \quad (3.15)$$

where  $\Delta N = \Delta N_{it} + \Delta N_{ox}$  (in unit  $cm^{-2}$ ),  $\mu_0$  is the original channel mobility,  $\alpha$  is a process dependent constant and  $\alpha \approx 2.4 \times 10^{-12} cm^2$  [56].

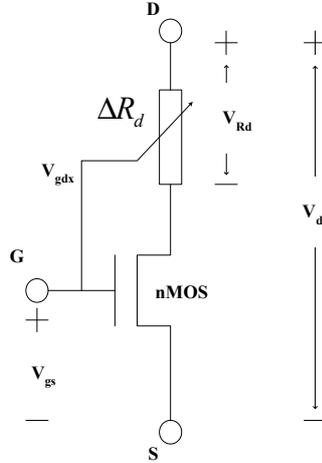


Figure 3.3: HCI circuit model in MaCRO. In the model:  $V_{gdx} = V_{gs} - V_t - V_{ds}$  and

$V_{Rd} = I_{ds} \Delta R_d$ .  $V_t$  is threshold voltage and  $I_{ds}$  is the current from node  $D$  to  $S$ .

The charge in conducting channel,  $Q_{ch}(y)$ , is modeled as:

$$Q_{ch}(y) = -C_{ox}(V_{gs} - V_t - \frac{q\Delta N}{C_{ox}} - V_{ch}(y)) \quad (3.16)$$

where  $C_{ox}$  is the gate oxide capacitance per unit area,  $V_{ch}(y)$  is the potential along the channel,  $y$  is the horizontal axis pointing to the drain and along the channel. All other parameters in equ.(3.16) assume their normal meaning.

Applying Gradual Channel Approximation (GCA) and combining equ.(3.15), the drain current,  $I_{ds}$ , with inclusion of hot carrier induced mobility degradation effect, is obtained as:

$$I_{ds} = \frac{\mu_0}{1 + \alpha\Delta N} C_{ox} \frac{W}{L} (V_{gs} - V_t - \frac{q\Delta N}{C_{ox}} - \frac{V_{ds}}{2}) V_{ds} \quad (3.17)$$

Now consider the circuit model in Fig.3.3 in which nMOS is the undamaged device with mobility  $\mu_0$  and threshold voltage  $V_t$ , and in series connection with  $\Delta R_d$ , the current from node  $D$  to  $S$  can be obtained as:

$$I_{ds} = \mu_0 C_{ox} \frac{W}{L} (V_{gs} - V_t - \frac{V_{ds} - V_{Rd}}{2}) (V_{ds} - V_{Rd}) \quad (3.18)$$

where  $V_{Rd}$  is voltage drop across  $\Delta R_d$ . Combining equ.(3.17) and equ.(3.18), and then solving for  $V_{Rd}$  yields:

$$V_{Rd} = -V_{gdx} + \sqrt{V_{gdx}^2 + 2V_{ds}\Delta N [\frac{\alpha(V_{gdx} + \frac{V_{ds}}{2})}{1 + \alpha\Delta N} + \frac{q}{C_{ox}}]} \quad (3.19)$$

where  $V_{gdx} = V_{gs} - V_t - V_{ds}$  for linear region and  $V_{gdx} = 0$  for saturation region.

According to equ.(3.17), when  $\Delta N = 0$  at  $t = 0$ , we get the undamaged drain current flowing through nMOS which is defined as  $I_{ds0}$ :

$$I_{ds0} = \mu_0 C_{ox} \frac{W}{L} (V_{gs} - V_t - \frac{V_{ds}}{2}) V_{ds} \quad (3.20)$$

If  $\Delta N$  is small, from equ.(3.17) and equ.(3.20), we can get a simple relation between fresh and degraded drain-to-source current:

$$I_{ds} = \frac{I_{ds0}}{1 + \alpha\Delta N} \quad (3.21)$$

Based on the above deduction, we finally obtain function of  $\Delta R_d$  which is determined by  $\Delta N$  and terminal voltages and currents:

$$\Delta R_d = \frac{1 + \alpha\Delta N}{I_{ds0}} V_{Rd} \quad (3.22)$$

where  $I_{ds0}$  is given by equ.(3.20) and  $V_{Rd}$  is given by equ.(3.19). In quasi-static operation,  $\Delta N$  is a time dependent parameter, therefore,  $\Delta R_d$  is also time dependent. At any time  $t$ , if  $\Delta N$  is known,  $\Delta R_d$  will be solely determined. The models for  $\Delta N_{it}$  and  $\Delta N_{ox}$  have been well documented in literature [45, 46].  $\Delta N_{it}$  can be obtained from equ.(3.1) if technology related constant  $C_1$  is extrapolated from device testing. The models and model parameters for  $\Delta N_{ox}$  are given in [46] (on pp.59-66). For convenience, they are recapitulated as follows.

The modeling of  $\Delta N_{ox}$  starts from a simple injection current model,  $I_{ei}$ , which describes one-dimensional process of electron injection into oxide based on quasi-elastic scattering assumption.

$$I_{ei} = \frac{1}{2} \frac{I_{ds}}{WL} \frac{t_{ox}}{\lambda_r} R^2 P_i(E_{ox}) \exp\left(-\frac{1}{R}\right) \quad (3.23)$$

where  $L$  is channel length,  $W$  is channel width,  $E_m$  is given by equ.(3.2),  $\lambda_r$  is re-direction mean-free path ( $\lambda_r = 61.6nm$ ),  $t_{ox}$  is oxide thickness.  $R = \lambda E_m / \varphi_b$ , where  $\lambda$  is the scattering mean-free path of the hot electron ( $\lambda = 9.2nm$ ), and  $\varphi_b$  denotes the silicon and oxide energy barrier ( $\varphi_b \approx 3.2eV$  for nMOSFET).

The most important term in equ.(3.23) is  $P_i(E_{ox})$ , which denotes the probability that a hot electron can enter the gate oxide by surmounting the surface potential barrier. An empirical expression for  $P_i(E_{ox})$  is given as:

$$P_i(E_{ox}) = \frac{\alpha E_{ox}}{1 + E_{ox}/\beta} \times \frac{1}{1 + \frac{\gamma}{L} \exp(-E_{ox}t_{ox}/1.5)} + \eta \quad (3.24)$$

where  $E_{ox} = (V_{gs} - V_{ds})/t_{ox}$ . Other model fitting parameters are given in [46] (on pp.62). Equ.(3.24) is for the case  $E_{ox} \geq 0$ , if  $E_{ox} < 0$ , it is simplified to  $P_i(E_{ox}) = \eta$ .

Based on equ.(3.23) and (3.24), for simulation purposes, a two-term kinetic equation is given in equ.(3.25) to model the relationship between oxide trapped charge density  $\Delta N_{ox}$  and electron injection current:

$$\Delta N_{ox} = N_1(1 - e^{-\sigma_1 I_{ei} t}) - N_2(1 - e^{-\sigma_2 I_{ei} t}) \quad (3.25)$$

A set of typical model fitting parameters for equ.(3.25) have been given in [46] (on pp.65).

The above new  $\Delta R_d$  model inherits all the merits of HISREM model and it is physically more comprehensive in characterizing hot carrier damages. The drawback of this improved  $\Delta R_d$  model is the inclusion of one more parameter  $\Delta N_{ox}$ , which complicates parameter extraction work.

For now, MaCRO does not provide pMOSFET HCI circuit model because HCI physical effects on pMOSFETs are weaker than those of nMOSFETs. With further scaling of CMOS devices, pMOSFET may suffer from more pronounced HCI damage than ever before. In future work, MaCRO will include pMOSFET HCI circuit model based on the channel shortening theory and SPICE macro models proposed in [57].

### 3.4 Implementation in MaCRO

With the models and equations presented in this chapter, the process of device and circuit HCI lifetime prediction and MaCRO reliability simulation with HCI circuit model are straightforward. With proper settings, SPICE simulator can predict  $I_{ds}$  for each nMOSFET from the fresh SPICE simulation with the original circuit schematic or netlist. Since  $I_{ds}$  is known,  $I_{sub}$  can be calculated from equ.(3.4). If  $A_{HCI}$  is set to 1 in equ.(3.9), then device normalized HCI lifetime  $t_f$  can be predicted. All these predicted lifetime values are then sorted from low to high to help identify the most degraded devices. Finally, HCI circuit model, i.e. Fig.3.3, is incorporated in the schematic or netlist to substitute these identified most degraded devices, and second round of SPICE simulation is performed to analyze the circuit reliability behaviors. The detailed simulation algorithm is given in the previous chapter. The Matlab program for calculating  $\Delta R_d$  values in terms of  $0.25\mu m$  technology parameters is shown in Appendix A.1. Because  $\Delta R_d$  of each identified nMOSFET is determined by its terminal voltage and current stresses up to the time  $t$ , designers can investigate circuit failure behaviors at any interested time and check circuit functionality which may be tampered by hot carrier effect. Moreover, designers can sweep  $\Delta R_d$  value in SPICE simulation and identify the critical value at which the circuit fails to operate as expected. The magnitude of this critical  $\Delta R_d$  value reflects circuit reliability margin because it is directly related to device hot carrier damage level. As a result, SPICE simulation with the  $\Delta R_d$  model provides circuit designers an expeditious way in evaluating performance and reliability

margins. The same process is also applied to TDDB and NBTI models which are discussed in the following two chapters.

### 3.5 Conclusion

The HCI lifetime model and circuit model (i.e. the new  $\Delta R_d$  model) are presented in this chapter. Substrate current and gate leakage current are major reliability monitors in device and circuit hot carrier lifetime predictions for nMOS-FET and pMOSFET, respectively. By lumping all common model parameters into a prefactor, device normalized lifetimes, rather than absolute lifetimes, can be quickly calculated and ranked to identify the most degraded transistors. The second round SPICE simulation with the inclusion of the new  $\Delta R_d$  model will reveal circuit performance degradation and reliability behavior under HCI stress at any given time. The new  $\Delta R_d$  model is improved to include the contribution of both interface trapped charge and oxide trapped charge, which is neglected in the original  $\Delta R_d$  model. This improvement, although complicates parameter extraction work, is physically more comprehensive and accurate in characterizing hot carrier damages.

## Chapter 4

### Time Dependent Dielectric Breakdown Effect and Models

#### 4.1 Introduction

Gate oxide breakdown is one of the most important failure mechanisms in CMOS technologies. When electric field is applied to the dielectric-isolated gate of a MOSFET, the progressive degradation of the dielectric material will result in the formation of conductive paths in oxide and short the anode and the cathode. When this happens, continuous stress of electric field on gate oxide may lead to excessive energy dissipated, or even thermal runaway, through breakdown paths. The electrical aftereffects of oxide breakdown are abrupt increase in gate current and loss of gate voltage controllability on device current flowing between drain and source. This kind of failure mechanism is known as Time Dependent Dielectric Breakdown (TDDB). The TDDB failures accelerate as the thickness of the gate oxide decreases with continued device scaling. ITRS'03 [58] predicts the equivalent oxide physical thickness for high-performance logic technology to be  $1.2nm$  at  $90nm$  technology node, however, oxides below  $2.5nm$  will not be able to sustain the operating voltage for their full expected lifetime [59]. Therefore, TDDB will become a potential reliability showstopper for sub- $100nm$  CMOS integrated circuits.

## 4.2 Accelerated Lifetime Model

TDDDB defect generation mechanism and device wearout dynamics have been extensively investigated in the past and there are many distinct, even controversial and contradicting, models having been proposed in literature. After many years of development, three successful models, thermochemical model, Anode Hole Injection (AHI) model and voltage driven model, are singled out and have gained broad applications.

Thermochemical model, also known as  $E$  model, assumes a direct correlation in existence between the electric field and the oxide degradation. The weak chemical bonds ( $Si - Si$  bonds) in  $SiO_2$  associated with oxygen vacancies experience heavy strains due to the high electric field applied across the oxide, and some bonds may obtain enough thermal energy to break off and create defects or traps which, when accumulated to large amount, will lead to oxide breakdown. According to thermochemical model, if the logarithm of time-to-failure  $t_f$  is plotted against applied electric field  $E$ , a straight line will be observed, therefore, lifetime can be modeled as:

$$t_f = B_1 \exp(-\gamma E_{ox}) \exp(E_a/\kappa T) \quad (4.1)$$

where  $E_{ox}$  is externally applied electric field across the dielectric in unit  $MV/cm$ ,  $\gamma$  is field acceleration factor (with typical value of 1.1 decade per  $MV/cm$  [60, 61]),  $E_a$  is the thermal activation energy ( $E_a = 0.6 \sim 0.9eV$  [62]), and  $B_1$  is technology constant. The  $E$  model has been proved to provide a good fit to data from long term low field TDDDB stresses.

AHI model assumes gate oxide breakdown is triggered by the trapping of holes at localized regions in oxide, which either enhances the cathode field or leads to the oxide electron trap generation, and increases the local current density. This facilitates local hole trapping and trap generation in a positive loop, and eventually leads to sudden breakdown of oxide [63]. Lifetime  $t_f$  function in an earlier version of AHI model derived a reciprocal electric field dependence ( $1/E$ ) from the functional form of Fowler-Nordheim (FN) electron tunneling current, which is the driving force for oxide defect generation, and impact ionization coefficient in  $SiO_2$ . In this case,  $t_f$  can be approximated as:

$$t_f = B_2 \exp(\beta/E_{ox}) \exp(E_a/\kappa T) \quad (4.2)$$

where  $\beta$  is the electric field acceleration factor (with typical value of  $350 MV/cm$ ),  $B_2$  is a process-dependent prefactor (the typical value is  $1 \times 10^{-11} s$  [62]). The  $1/E$  model has been proved to provide a good fit to data from long term high field TDDB stresses. It is important to note that AHI model does not predict a strict  $1/E$  dependence [63], and there exists a model which predicts a much stronger  $1/E$  effect ( $t_f \propto \exp(\beta/E_{ox})(1/E_{ox}^2)$  [64]).

Each of the two models ( $E$  and  $1/E$ ) can only fit data in a limited range of electric field, which may lead to significant errors in lifetime extrapolation if one exclusively uses only one of them in reliability analysis. Researchers have proposed parallel competing models (i.e. combined models) in terms of  $E$  and  $1/E$  models trying to account for TDDB data in a larger electric field range [63, 65].

The applicability of  $E$  and  $1/E$  models is mainly valid for oxides thicker than

5nm where non-ballistic electron injection due to FN tunneling is dominant. When gate oxide thickness is smaller than 5nm (i.e. ultrathin oxide), gate oxide lifetime dramatically shortens with the increase in direct tunneling current. In this situation, the validity of electric field driven models becomes problematic because the injected electrons will travel ballistically through oxide without entering oxide conduction band and the electron energy at the anode is controlled by the applied gate voltage [66]. This new phenomenon of electron injection in ultrathin oxides prompts to the generation of voltage driven breakdown models, in which the dependence of lifetime  $t_f$  on gate voltage  $V_{gs}$  is given by [67, 68]:

$$t_f = B_3 \exp(-\theta V_{gs}) \exp(E_a/\kappa T) \quad (4.3)$$

where  $\theta$  is voltage acceleration factor,  $B_3$  is technology constant. The typical values of  $\theta$  and activation energy  $E_a$  are given in [67, 69].

All the TDDB lifetime models presented so far are based on exponential law for field or voltage acceleration and Arrhenius law for temperature acceleration. Recent work shows that these two acceleration laws may be not accurate as gate oxide thickness scales below 5nm, and the extrapolation of ultrathin oxide lifetime with these exponential relations may produce erroneous or even absurd results. According to experimental data, the exponential law for time-to-breakdown voltage dependence cannot hold over a wide range of gate voltage, otherwise, the extrapolation of lifetime down to normal use conditions will predict (1) the lifetime of smaller-area structures would be shorter than that of larger-area structures, and (2) the lifetime of thinner oxide devices would ultimately exceed that of thicker oxide device, both of which

contradict oxide degradation physics [70]. Therefore, new TDDB acceleration laws for voltage and temperature must be explored in conjunction with CMOS technology development.

Voltage and temperature acceleration laws of oxide breakdown and their inter-relationship are critical factors for understanding ultrathin oxide reliability. Recent experimental data shows oxide time-to-breakdown evolution with temperature does not exactly follow an Arrhenius law: the activation energy increases with temperature. This behavior may be explained either by non-thermochemical origin for the breakdown mechanism or by a competing model involving two distinct mechanisms with different activation energies [68]. As to the oxide time-to-breakdown evolution with voltage, Wu et al at IBM [71, 72] proved with convincing data that the voltage dependence of time-to-breakdown follows a power law behavior rather than an exponential law as commonly assumed. The ultrathin oxide power law dependence of lifetime on gate voltage is consistent with the experimental observations that voltage exponential law acceleration factor  $\theta$  (shown in equ.(4.3) and defined as  $\theta = -\partial \ln t_f / \partial V_{gs}$ ) is (1) temperature dependent at a fixed gate voltage, and (2) voltage dependent at a fixed temperature. Due to these new oxide time-to-breakdown voltage and temperature dependencies and the complicated interaction between voltage and temperature, TDDB lifetime modeling becomes much more difficult than ever before. In another perspective, however, the power law voltage dependence and non-Arrhenius temperature acceleration provide possible relief in circuit reliability margin which has quickly diminished due to the scaling of oxide thickness [73].

Ultrathin oxide lifetime model in MaCRO is similar to the model proposed by Wu et al at IBM [70, 72] with some improvements including the addition of oxide Poisson area scaling statistics and cumulative failure percentile scaling law. The original Wu model (i.e. power law voltage acceleration and non-Arrhenius temperature acceleration) has been implemented in the Reliability Aware Micro-Processor (RAMP) model jointly developed by UIUC and IBM for long-term processor reliability prediction [74, 75].

On the basis of extensive experimental investigation, ultrathin oxide lifetime dependence on voltage (power law acceleration) can be accurately captured by two simple empirical formulae [70, 72]:

$$\frac{V_{gs}}{t_f} \frac{\partial t_f}{\partial V_{gs}} = n(T) \quad (4.4)$$

and

$$\frac{d}{dT} \left( \frac{1}{t_f} \frac{\partial t_f}{\partial V_{gs}} \right) \Big|_{t_f(\%)} = 0 \quad (4.5)$$

where  $n(T)$  denotes the temperature dependent voltage acceleration factor, and  $T$  is absolute temperature in Kelvin,  $t_f(\%)$  means lifetime for a fixed cumulative percentile of failure (for example 63%).

Equ.(4.4) reflects the power law dependence of time-to-breakdown on voltage: if  $t_f = t_0 V^{n(T)}$ , then  $\partial t_f / \partial V = n(T) t_0 V^{n(T)-1} = n(T) t_f / V$ , so  $(V/t_f)(\partial t_f / \partial V) = n(T)$ . Equ.(4.4) also reflects the experimental fact that  $\theta$  ( $\theta = -\partial \ln t_f / \partial V$ ) is a voltage dependent voltage acceleration factor:  $\theta = -n(T)/V$ . Equ.(4.4) shows that voltage power law acceleration factor  $n(T)$  is temperature dependent, and for simplicity we assume a linear relation  $n(T) = a + bT$  (note:  $n(T)$  should be always

less than 0). This leads to the first part of TDDB lifetime model equation in MaCRO:

$$t_f \propto V_{gs}^{a+bT} \quad (4.6)$$

Equ.(4.5) reflects the experimental fact that at a fixed accumulative failure percentile lifetime, voltage exponential law acceleration factor  $\theta$  is temperature independent. In reliability test, engineers normally stress a large number of samples to a high cumulative percentile of failure (e.g.  $F = 63\%$ ) and calculate lifetime at this percentile (e.g.  $t_f(63\%)$ ), then they extrapolate lifetime to a low cumulative percentile of failure (e.g.  $F = 0.01\%$ ) at normal use condition. In order to take into account the effect that different cumulative failure percentiles may be selected in use conditions for different devices even though they are tested to the same high cumulative failure percentile, it is necessary to incorporate Weibull statistics of oxide breakdown in the lifetime model development.

According to Weibull distribution, the cumulative failure probability  $F(t)$  is:

$$F(t) = 1 - \exp[-(t/\alpha)^\beta] \quad (4.7)$$

where  $\alpha$  is the characteristic life (i.e. lifetime at 63%) and  $\beta$  is the slope parameter which represents trends of failure rate. Weibull distribution is an extreme-value distribution in  $\ln(t)$  and can model weakest-link type of failure mechanisms. TDDB is a weakest-link mechanism because the first breakdown of any small portion in the gate oxide of a device will lead to the failure of the device and the whole circuit [76].

Equ.(4.7) can be rearranged and modified to:

$$t_f = \alpha \left[ \ln \frac{1}{1-F} \right]^{\frac{1}{\beta}} \quad (4.8)$$

At normal use conditions, lifetime is often defined as the time to a very small cumulative percentile of failure (e.g.  $F = 0.01\%$ ), therefore, applying logarithmic approximation law on equ.(4.8), we obtain the second part of TDDB lifetime model equation in MaCRO:

$$t_f \propto F^{\frac{1}{\beta}} \quad (4.9)$$

Another effect needs to consider in TDDB lifetime model is that the gate oxide areas of sampled devices in accelerated tests are normally significantly different from those of devices in circuits. Experimental observations prove that the lifetime of TDDB is a function of the total gate oxide surface area due to the weakest-link character of oxide breakdown [70]. This gate oxide area scaling effect has been modeled in [70, 73, 76], and leads to the third part of TDDB lifetime model equation in MaCRO:

$$t_f \propto \left(\frac{1}{WL}\right)^{\frac{1}{\beta}} \quad (4.10)$$

where  $W$  is the channel width and  $L$  is the channel length.

Finally, for the temperature acceleration effect, a non-Arrhenius model has been proposed in [70, 72], which is the fourth part of TDDB lifetime model equation in MaCRO:

$$t_f \propto \exp\left(\frac{c}{T} + \frac{d}{T^2}\right) \quad (4.11)$$

where  $c$  and  $d$  are voltage dependent constants. In equ.(4.11), the second term  $d/T^2$  empirically inserts non-Arrhenius temperature effects in the lifetime model.

Combining equ.(4.6), (4.9), (4.10) and (4.11), we obtain a complete TDDB lifetime model for ultrathin oxides:

$$t_f = A_{TDDB} \left(\frac{1}{A}\right)^{\frac{1}{\beta}} F^{\frac{1}{\beta}} V_{gs}^{a+bT} \exp\left(\frac{c}{T} + \frac{d}{T^2}\right) \quad (4.12)$$

where  $A = W \times L$  is the device gate oxide area,  $\beta$  is Weibull slope parameter,  $F$  is cumulative failure percentile at use condition (assuming the same cumulative failure percentile at test conditions),  $V_{gs}$  is gate-to-source voltage,  $T$  is temperature,  $a$ ,  $b$ ,  $c$ ,  $d$  and  $A_{TDDB}$  are model fitting parameters determined from experimental work. A set of typical values of these parameters are:  $\beta = 1.64$ ,  $F = 0.01\%$ ,  $a = -78$ ,  $b = 0.081$ ,  $c = 8.81 \times 10^3$  and  $d = -7.75 \times 10^5$  [74, 75].

It is important to note that equ.(4.12) is most applicable to the cases when the gate oxide thickness is thinner than  $5nm$  (corresponding to  $0.25\mu m$  technology and beyond). If the gate oxide thickness is larger than  $5nm$ , in order to simplify parameter extrapolation work, equ.(4.3) should be used instead with the default value of  $\theta$  as 32. If the gate oxide thickness is much larger than  $10nm$ ,  $E$  or  $1/E$  model (equ.(4.1) and (4.2), respectively) should be used depending on the magnitude of power supply voltage.

### 4.3 Failure Equivalent Circuit Model

It is an onerous work to develop an effective circuit model for gate oxide breakdown because device post-breakdown behaviors are extremely complicated, sometimes even perplexing. Device  $I-V$  characteristics after gate oxide breakdown relies on many parameters including breakdown location, transistor type, voltage

polarity, device operation mode (accumulation or inversion), oxide area and even poly-gate doping type. Nevertheless, literature review reveals an interesting phenomenon that TDDB failure circuit modeling is a very active area and more than a dozen of circuit models have been developed by various research institutes and industrial labs. All this work attempts to develop quantitative methodologies for predicting the response of circuits to device's gate oxide breakdown events [77]. In this section, first, some of the most successful TDDB failure circuit models are reviewed, which is necessary because of the obvious absence of papers of this kind in literature, the introduction of the TDDB circuit model adopted in MaCRO is given afterwards.

Starting from the observation that a CMOS inverter's transfer curve under gate oxide stresses can be fitted by a combination of a threshold voltage shift (caused by charge trapping prior to breakdown) and a gate-to-drain leakage current model which follows the form of a power-law relation as  $I = KV_{gd}^p$ , Rodriguez et al at IBM [78, 79, 80] developed a simple TDDB circuit model which consists of two voltage-dependent current sources bridging gate-to-drain and gate-to-source, respectively, allowing the oxide breakdown leakage current in a transistor to be simulated in a circuit. This power-law leakage current model is illustrated in Fig.4.1.

The effects of gate oxide breakdown on the stability of SRAM cells and ring oscillators have been analyzed with this power-law leakage current model. Results show that for SRAM cells, oxide breakdown at different locations (drain, p-source and n-source) leads to different trends in noise margin degradation, while for ring oscillators, oxide breakdown changes the loading of neighboring inverter stages and

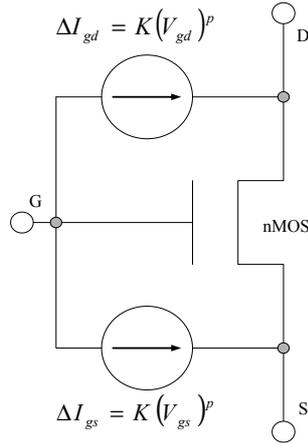


Figure 4.1: Power-law leakage current model. The exponent  $p$  varies from 5 to 2 as the degradation level increases.  $K$  reflects the “size” of the breakdown spot.

degrades the VTC [78].

Rodriguez et al [80] noted that a linear ohmic oxide breakdown resistance is not sufficient to model the experimental data. The ohmic model only provides good results for hard breakdown, but the power-law leakage current model predicts progressive oxide breakdown behaviors much better prior to the final hard breakdown.

In a MOSFET, the oxide breakdown changes isolations of the device’s internal structures by forming an abnormal conduction path and this effect can be modeled with parasitic ohmic or rectifying device elements depending on the relative doping of the internal structures being shorted. Based on the fact that oxide post-breakdown behavior depends on breakdown location (gate-to-channel, gate-to-drain and gate-to-source), transistor type (nMOSFET and pMOSFET) and poly-gate doping type ( $n^+$  poly-gate and  $p^+$  poly-gate), Segura et al [81, 82] developed a complete set of Gate Oxide Short (GOS) electrical models (altogether 12 different GOS models)

to account for all combinations of these location and doping effects. Among these models, the most important one is the model for gate-to-channel breakdown of nMOSFET with  $n^+$  poly-gate. For this kind device, the gate-to-channel breakdown path between  $n^+$  poly-gate and  $n$  type inversion channel can be modeled as a gate-to-channel resistance  $R_{GOS}$ . The formation of this resistance-like breakdown path splits the whole channel into two parts, which is physically equivalent to two transistors connected in series. This model is illustrated in Fig.4.2.

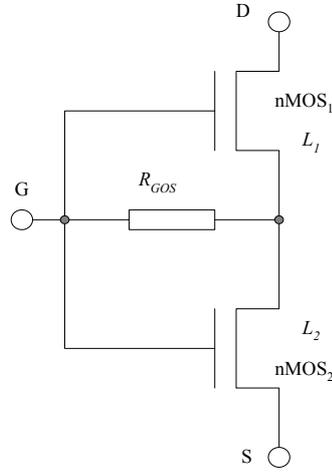


Figure 4.2: TDDB GOS model for gate-to-channel breakdown of nMOSFET with  $n^+$ -poly gate. The channel lengths of nMOS1 and nMOS2 follow the relation:  $L_1 + L_2 = L$  where  $L$  is the undamaged nMOSFET channel length. The parameter  $R_{GOS}$  is related to the size and location of the breakdown path. A value of  $R_{GOS}$  as low as  $3K\Omega$  was used in the simulation in [82].

For other combinations of location and doping effects, the models can be readily deduced with the similar principle. For example, when the breakdown path appeared between the gate and the drain (or the source) terminals of the nMOSFET,

an  $n^{++} - n^+$  barrier (i.e.  $n^+$  poly-gate to  $n^+$  drain/source diffusion) will form. In this case the breakdown is modeled with a resistance between gate-to-drain/source.

With these GOS electrical models, Segura et al [81] explored testing considerations at circuit level to sensitize GOS under various logic fault situations (stuck-at, stuck-open and stuck-on faults) and concluded that GOS does not behave as a bridge in normal cases and stuck-at based Automatic Test Pattern Generation (ATPG) may not detect GOS depending on the gate topology.

Gate oxide breakdown equivalent circuit models for analog circuits and RF circuits are also developed in an attempt to expand model applicability and explore oxide breakdown effect beyond digital circuits. For typical analog circuits, oxide breakdown changes parameters of transistors in differential pairs in an asynchronous way and therefore leads to mismatches, which accelerates the offset generation and compromises circuit functionality [83]. As for RF circuits, they are very sensitive to device parameter's drift, therefore, oxide breakdown is expected to have more severe impact on their functionality and performance [84].

Yang et al [84, 85] developed an RF failure circuit model for gate oxide breakdown and investigated the effect of TDDB on a Low Noise Amplifier (LNA) circuit. This RF equivalent model is shown in Fig.4.3 which consists of the original nMOS-FET, the terminal series resistances ( $R_G$ ,  $R_D$ ,  $R_S$ ), the substrate parasitic resistances ( $R_{DB}$ ,  $R_{SB}$ ,  $R_{DSB}$ ), gate overlap parasitic capacitances ( $C_{GDO}$ ,  $C_{GSO}$ ), the junction capacitances ( $C_{jDB}$ ,  $C_{jSB}$ ), and the two inter-terminal resistances ( $R_{GD}$ ,  $R_{GS}$ ).  $R_G$  and the "H" type substrate  $RC$  network are included for more accurate RF modeling. The two resistances  $R_{GD}$  and  $R_{GS}$  vary in opposite directions rep-

representing different breakdown locations along the channel from source to drain. If one of them is significantly smaller than the other, breakdown is gate-to-source or gate-to-drain depending on which resistance is dominant.

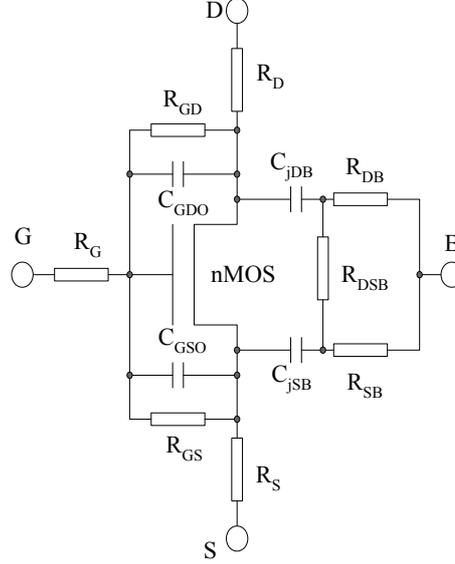


Figure 4.3: TDDB RF equivalent circuit model. Model parameters for simulation in [84] are set as:  $R_G = 85.4\Omega$ ,  $R_D = R_S = 12.14\Omega$ ,  $R_{GD} = 6.88K\Omega$ ,  $R_{GS} = 23K\Omega$ ,  $C_{GDO} = C_{GSO} = 15.3fF$ ,  $C_{jDB} = C_{jSB} = 7fF$ ,  $R_{DSB} = 80K\Omega$ ,  $R_{DB} = R_{SB} = 49.37\Omega$ .

Based on this TDDB RF circuit model, the performance degradation of  $0.16\mu m$  nMOSFET devices and a  $1.8GHz$  LNA circuit is analyzed [84]. For the device  $S$ -parameters, the inclusion of  $R_{GD}$  and  $R_{GS}$  changes device input impedance  $S_{11}$ , provides an additional connection between gate and drain and therefore degrades reverse transmission coefficient  $S_{12}$ , changes the output impedance  $S_{22}$  at the drain, and also decreases transconductance  $g_m$  which is equivalent to forward transmission

coefficient  $S_{21}$ . For the LNA circuit, oxide breakdown has significant impact on its performance: most  $S$ -parameters drift dramatically and fail to meet usual performance requirements, input impedance matching is disturbed due to increased gate leakage current, and noise figure obviously deteriorates with the breakdown path forming across the gate oxide, which adds another noise source to the transistor.

Up to now, the most frequently discussed TDDB circuit model is the one proposed by Kaczer et al at IMEC [86, 87, 88, 89, 90, 91]. In this model, the breakdown path is assumed to be formed by  $n$ -type silicon and a microscopic structure of the device is explored to investigate the exact configuration and connection of device internal parts after gate oxide breakdown. For an nMOSFET ( $n^+$  poly-gate/ $p$  substrate/ $n^+$  drain and source diffusion) with an oxide breakdown path formed between gate and substrate, if the gate voltage is negative ( $V_G < 0$ ), the device is in accumulation state and no inversion layer is developed below the  $Si - Si_2$  interface. The contact region of the breakdown path ( $n$ -type) and the substrate ( $p$ -type) is a forward biased  $pn$  junction. Electrons emit from  $n^+$  poly-gate, flow through  $n$ -type breakdown path, diffuse along the substrate and are collected by the source and the drain junctions. This mechanism is exactly that of a bipolar transistor with emitter at the breakdown path, base at the substrate and collector at the source and the drain. Therefore, nMOSFET with oxide breakdown and operated at negative gate voltage can be modeled with a gate resistor, two bipolar transistors and the original nMOSFET [86, 88]. Because nMOSFETs rarely operate in negative gate voltage bias situation, this complicated two-bipolar-transistor model for ( $V_G < 0$ ) is not of primary interest.

When gate voltage is positive enough such that nMOSFET is in strong inversion state, an  $n$ -type conduction channel will form under the gate oxide connecting the source and the drain. Now the contact region of the breakdown path ( $n$ -type) and the channel ( $n$ -type) is an ohmic connection. The positive gate voltage forces electric field penetrate through the breakdown path and deplete the contact region of breakdown path and substrate. This contact region serves as electron sink and therefore it can be treated as an additional drain in the middle of the channel. Based on this microscopic picture, an equivalent electrical circuit for nMOSFET with hard gate oxide breakdown and operated in positive gate voltage is constructed and illustrated in the Fig.4.4.

Apart from the original nMOSFET ( $nMOS$ ), the model contains a constant resistance ( $R_G$ ) corresponding to breakdown path, two adjacent parasitic nMOSFETs ( $M_S$  and  $M_D$ , characterized by level-1 SPICE models), and two resistors ( $R_S$  and  $R_D$ ) characterizing the resistance in the source and the drain extensions, respectively. The effects of breakdown location are represented by varying the gate lengths of  $M_S$  and  $M_D$ . Gate-to-channel breakdowns in the vicinity of the drain or the source are represented by logarithmically varying extension resistances  $R_S$  or  $R_D$  [86]. For gate-to-source (or gate-to-drain) breakdowns, the model can be simplified to a circuit containing only  $R_G$ ,  $R_S$  (or  $R_D$ ) and the original  $nMOS$  transistor.

This model has been used in a CMOS ring oscillator oxide breakdown analysis [87]. The simulation shows that gate-to-channel breakdowns have minor effect on circuit operation but breakdowns at the very edges of the gate significantly damage the circuit performance. This observation reveals that progressive breakdown (i.e.

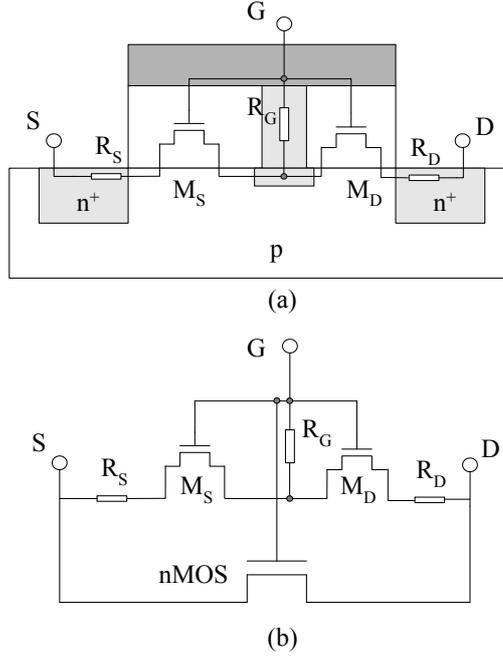


Figure 4.4: TDDDB circuit model for nMOSFET with hard gate oxide breakdown and operated in positive gate voltage. (a) Cross-sectional view of breakdown structure. (b) Equivalent circuit model. Model parameters for simulation in [86] are set as:  $R_G = 1K\Omega$ ,  $L_{M_S} + L_{M_D} = 0.09\mu m$ ,  $W_{M_S} = W_{M_D} = 0.25\mu m$ ,  $R_D$  and  $R_S$  vary from  $2.5K\Omega$  (at source and drain) to  $12.5K\Omega$  (at the middle of the channel).

soft breakdown) occurs mainly in the transistor channel, while hardest circuit-killing breakdowns occur above the source and the drain extension regions [89]. This conclusion can be explain with the help of the Kaczer model: in the extension regions where contact resistances are low, the power dissipation during the breakdown is very high and leads to accelerated wearout of the breakdown path, this corresponds to hard breakdown behaviors; if breakdown happens in the transistor channel region, where resistance (i.e. channel resistance) of the discharge path is higher, soft

breakdown will be triggered.

Even though a lot of work has been done to mature this model, careful evaluations in [84] and our critical examinations have identified several limitations of this Kaczer model: (1) The level-1  $M_S$  and  $M_D$  models are obsolete. (2) The model only applies to linear operation situation, if breakdown path forms above the saturation region where channel has “pinched-off”, the inclusion of the two parasitic transistors i.e.  $M_S$  and  $M_D$ , is not valid. (3)  $M_S$  and  $M_D$  bring two more drain diffusion regions, which do not physically exist. (4) Simulator cannot handle the breakdown position from zero to the whole channel length. (5) It is problematic to preserve the original nMOSFET in the model if  $M_S$  and  $M_D$  are included because they already represent all device internal structures after oxide breakdown. Specifically, the whole conducting channel has been physically characterized by  $M_S$  and  $M_D$ , therefore, it is erroneous to keep the original nMOSFET in the post-breakdown TDDB circuit model. (6) The prime assumption that the breakdown path is  $n$ -type silicon is arbitrary and not physically justified. The last two points are the most important ones and they prompted to develop a physically justifiable circuit model for gate oxide breakdown.

Besides what have been briefly reviewed above, there are many other successful models worth mentioning [92, 93, 94, 95, 96, 97]. A pMOSFET gate-to-channel breakdown model is proposed in [92] and used to investigate its effect on logic gate failures. A pair of breakdown models for nMOSFET and pMOSFET (only gate-to-diffusion breakdowns) is proposed and used to transform the effect of oxide breakdown into a delay fault or a logic fault [93]. Yeoh et al [94, 95] conducted a

thorough investigation of oxide breakdown modes and developed a set of complex models by combining resistors, diodes and transistors in different ways to model device internal connections after oxide breakdown path formed at different locations. Based on the work of linear non-split MOS model and non-linear two-dimensional channel split MOS model [96], a non-linear non-split MOS oxide breakdown model is developed in [97] in an attempt to enable circuit simulation of gate-to-channel effect on minimum length transistors. Even though these models do not accurately model all aspects of breakdown, the development of fundamental concepts, physical principles and modeling techniques in these models is the foundation work for constructing any advanced oxide breakdown circuit models. Following this conclusion, a new TDDB circuit model adopted in MaCRO is developed below.

From semiconductor materials point of view, it is improper to assume the breakdown path as  $n$ -type silicon diffusion because this is not physically substantiated, and the oxide breakdown path is actually defect-assisted electron conduction rather than a reliable physical connection. Therefore the resistance cannot be solely used to model gate-to-channel and gate-to-diffusion breakdowns. The correct modeling method should base on the channel potential re-distribution concept. Oxide breakdown path disturbs device channel surface potential in the vicinity below the breakdown path, where GCA assumption is broken, so new three-dimensional channel potential model has to be developed for this purpose. According to [82], if defining a three-dimensional coordinate system in terms of the gate oxide surface with  $x$  along the channel length  $L$  direction from source to drain,  $y$  perpendicular to the gate oxide, and  $z$  along the channel width  $W$  direction, then the contact point

of the breakdown path to channel surface can be defined as:  $x = L_1$ ,  $y = 0$  and  $z = W_1$  (refer to Fig. 10 in [82]). The drain current  $I_D$  of a defect-free MOSFET can be obtained from:

$$I_D = \frac{W}{L} [f(\Psi(x = L)) - f(\Psi(x = 0))] \quad (4.13)$$

where  $\Psi(x)$  is the channel surface potential at  $x$ ,  $f$  is a function of channel mobility, oxide capacitance, threshold voltage and device terminal voltages.

If the breakdown defect located at ( $x = L_1$ ,  $y = 0$  and  $z = W_1$ ) is considered, the two-dimensional channel can be divided into two regions, and similar to equ.(4.13), the drain and source currents of the damaged MOSFET can be written as [82]:

$$I_D = \frac{W}{L - L_1} [f(\Psi(x = L)) - f(\Psi(x = L_1))] \quad (4.14)$$

and

$$I_D = \frac{W}{L_1} [f(\Psi(x = L_1)) - f(\Psi(x = 0))] \quad (4.15)$$

where  $\Psi(x = L_1)$  is the surface potential under the breakdown path. Equ.(4.14) and (4.15) show that an nMOSFET with gate oxide breakdown is equivalent to the series connection of two devices with gate geometries of  $(W, L_1)$  and  $(W, L - L_1)$ .

No matter what the breakdown path is made of, its electrical effect is that it provides a conduction path to inject electrons from channel into gate, therefore a voltage dependent current source  $I_{OX}$  connecting between gate and channel can be used to model this effect. Based on the above discussion, a new TDDB circuit model is obtained and illustrated in Fig.4.5.

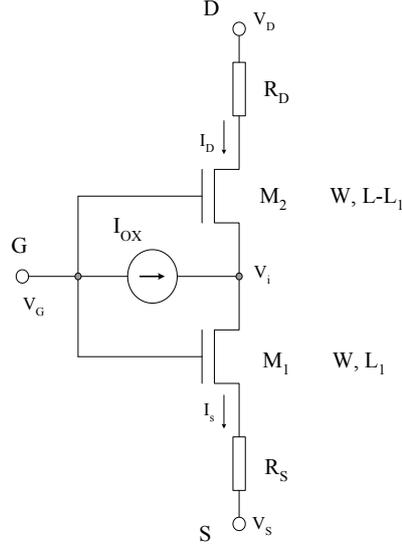


Figure 4.5: MaCRO TDDDB circuit model for nMOSFET with hard gate oxide breakdown.  $I_{OX} = I_S - I_D$  is a voltage dependent current source representing breakdown path current injection effect.  $R_D$  and  $R_S$  characterize the resistance in the source and the drain extensions, respectively.  $L_1$  represents breakdown location away from the source edge.

It seems this model requires two model parameters ( $L_1$  and  $V_i$ , which is voltage at the connection point of  $M_1$  and  $M_2$ ), but with some practical simplifications,  $V_i$  can be reduced to a function dependent on  $L_1$ . Therefore, there is only one independent model parameter left and requiring characterization, which facilitates the application of this model.

Suppose the original drain-to-source current of a fresh nMOSFET is  $I_{DS0}$ , and neglect effect of  $R_D$ ,  $R_S$  and short-channel effect (in order to simplify equation derivation), we can write  $I_{DS0}$  as:

$$I_{DS0} = \mu_n C_{ox} \frac{W}{L} [(V_{GS} - V_t)V_{DS} - \frac{1}{2}V_{DS}^2] \quad (4.16)$$

Apply Kirchhoff's Current Law (KCL) to Fig.4.5 (for simplicity, neglect  $R_D$  and  $R_S$ ), we can get the following equations:

$$I_{OX} = I_S - I_D \quad (4.17)$$

$$I_D = \mu_n C_{ox} \frac{W}{L_2} [(V'_G - V_i)(V_D - V_i) - \frac{1}{2}(V_D - V_i)^2] \quad (4.18)$$

$$I_S = \mu_n C_{ox} \frac{W}{L_1} [(V_{GS} - V_t)(V_i - V_S) - \frac{1}{2}(V_i - V_S)^2] \quad (4.19)$$

where  $L_2 = L - L_1$  is the channel length of  $M_2$ ,  $V'_G = V_G - V_{t2}$ ,  $V_{t2}$  is the original threshold voltage  $V_t$  plus body bias ( $V_{sb} = V_i$ ) induced enhancement effect.  $V_i$  represents the channel potential at the breakdown location.

The main effects of gate oxide breakdown on device characteristics are abrupt gate current and substrate current generation, and gate voltage cannot control and sustain channel current as strong as before, which leads to degradation of drain current. Therefore, a good assumption in Fig.4.5 is the source current  $I_S$  maintains its value as before, whereas injection of  $I_{OX}$  degrades  $I_D$  current at the drain. This means  $I_S = I_{DS0}$ . So from equ.(4.16) and equ.(4.19), we can solve for  $V_i$ :

$$V_i = V_{G_{on}} - \sqrt{V_{G_{on}}'^2 - (V_S^2 + 2V_{ov}V_S + \frac{2I_{DS0}L_1}{\mu_n C_{ox}W})} \quad (4.20)$$

where  $V_{G_{on}} = V_G - V_t$ ,  $V_{ov} = V_{G_{on}} - V_S$  is the gate overdrive voltage. If  $V_S$  is tied to ground, equ.(4.20) is reduced to:

$$V_i = (V_G - V_t) - \sqrt{(V_G - V_t)^2 - \frac{2I_{DS0}L_1}{\mu_n C_{ox}W}} \quad (4.21)$$

Equ.(4.21) (or equ.(4.20) if  $V_S \neq 0$ ) shows that  $V_i$  is solely determined by  $L_1$ . Therefore, the number of model parameters is reduced from two to only one. If the

breakdown location parameter  $L_1$  is characterized from experimental work, from equ.(4.16) ~ (4.21), the voltage dependent current source  $I_{OX}$  can be obtained.

The above nMOSFET TDDB circuit model can be easily extended to pMOS-FET by properly changing current flowing directions in Fig.4.5 and voltage/current signs in model equations.

#### 4.4 Implementation in MaCRO

Circuit designers may be baffled when they first come across  $L_1$  parameter and try to characterize it, but the systematic work conducted in [86, 87, 88, 89, 90, 91] provides insight on how to deal with  $L_1$  in circuit reliability simulation. If effective gate resistance after gate breakdown is defined as  $R_G$  ( $R_G = V_G/I_G$ ), according to [87], the relation between  $R_G$  and breakdown location  $L_1$  can be illustrated as Fig.4.6. It is clearly shown that effect of oxide breakdown is rather insensitive to breakdown locations if  $L_1$  is far away from the drain and source edges. This prompts to a practical simplification that for gate-to-channel breakdown, it is unnecessary to accurately determine  $L_1$  and this will not incur intolerable errors in simulation. The reason for this insensitivity is as follows: if  $L_1$  increases, then  $V_i$  also increases (by equ.(4.21)), but in the middle range of the channel ( $0 \ll L_1 \ll L$ ),  $I_D$  and  $I_S$  do not change monotonically with  $V_i$ , their combined effect maintains a roughly constant  $I_{OX}$  level. This effective gate resistance insensitivity is valid for  $L_1$  in the wide middle part of the channel. When  $L_1$  is very close to the ends of channel,  $I_D$  and  $I_S$  will change dramatically which corresponds to the abrupt decreases in

effective gate resistance  $R_G$  at both ends of the channel as indicated in Fig.4.6.

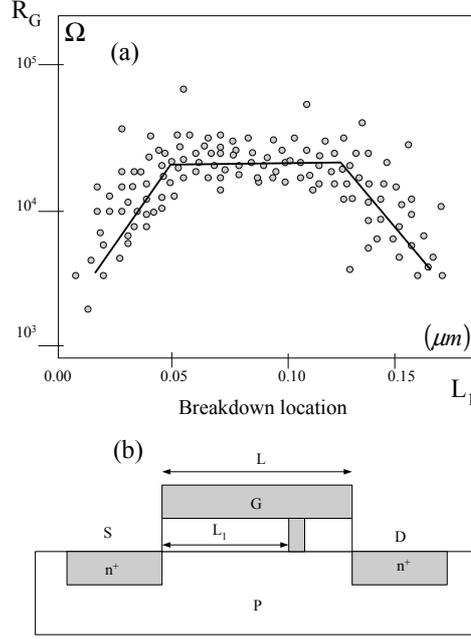


Figure 4.6: Relationship between effective gate resistance  $R_G$  and oxide breakdown location  $L_1$ . The bias condition is  $V_G = 1.5V$ ,  $V_D = V_S = 0$ . (a) Reproduce from [87] with illustrative data. (b) Cross-sectional view of breakdown location.

Based on the above discussion, in MaCRO simulation aiming at a quick investigation of gate-to-channel breakdown on circuit functionality, designers can arbitrarily select an  $L_1$  value (for example  $L_1 = 1/2L$ ) and calculate  $I_{OX}$  by device terminal voltage and current waveforms.  $R_D$  and  $R_S$  can be included to account for series resistance effects due to source/drain extensions. Their typical values for gate-to-channel breakdown are  $R_D = R_S = 12.5K\Omega$ . The Matlab program for calculating  $I_{OX}$  values in terms of  $0.25\mu m$  technology parameters is attached in Appendix A.2. It is worth emphasizing again that body effect of  $M_2$  in Fig.4.5 must be considered,

otherwise calculations would produce erroneous results.

The model shown in Fig.4.5 is not valid if  $L_1$  approaches to upper or lower boundaries. When  $L_1$  approaches 0 or  $L$ ,  $I_{OX}$  will bypass or sustain all current flowing through the transistor  $M_2$  or  $M_1$ , this is not physically correct, and SPICE circuit simulator also cannot handle the situation where a transistor's channel length is approaching 0. Therefore, in the case of gate-to-diffusion breakdowns, the two-transistor model collapses into one-resistor-shortening model (gate-to-source resistor  $R_{GS}$  or gate-to-drain resistor  $R_{GD}$ ). In MaCRO, the typical values of these resistances are set to:  $R_{GD} = R_{GS} = 2.5K\Omega$ . Even though gate-to-diffusion breakdown has much severe impact on circuit functionality, from statistics point of view, these events should rarely happen (they only occupy very limited portion of the horizontal axis in Fig.4.6 and most breakdown data crowd in the middle region). Therefore, in typical circuit reliability simulation and analysis, designers only need to focus on gate-to-channel breakdown effects, which is in contrast to most other work whose main concern is on gate-to-diffusion breakdown effect. Since gate-to-channel breakdown has less damage effect on device operation than gate-to-diffusion breakdown, MaCRO reliability simulation generally predicts relaxed TDDB effects on circuit functionality. This point and the resultant benefits are proved in Chapter 8 during SRAM reliability simulation and analysis.

## 4.5 Conclusion

A set of new TDDB lifetime model and circuit model are presented in this chapter. The lifetime model unifies many important experimental observations of oxide breakdown behaviors including power law voltage acceleration, non-Arrhenius temperature acceleration, Poisson area scaling statistics and cumulative failure percentile scaling effect. A thorough overview of existing TDDB circuit models is presented to compensate for the obvious absence of review papers of this kind in this area. Some limitations and an error in the most frequently discussed TDDB circuit model proposed by IMEC are identified. Finally, a new TDDB circuit model is developed and the number of model parameters is reduced to only one, which significantly simplifies the application process of this new failure model in circuit reliability simulation and analysis.

## Chapter 5

### Negative Bias Temperature Instability Effect and Models

#### 5.1 Introduction

Negative Bias Temperature Instability (NBTI) is a relatively new MOSFET intrinsic wearout mechanism which mainly occurs in pMOSFETs when they are stressed with negative gate voltage at elevated temperature. The typical stress conditions for NBTI effects are temperatures in the range of  $100 \sim 250^\circ C$  and oxide electric fields below  $6MV/cm$  which are smaller than those capable of initiating HCI effects. Therefore, NBTI is more severe than HCI for ultrathin oxides at low electric fields. Either negative gate voltages or elevated temperatures can induce NBTI, but a much severe degradation effect will be produced by their combination and interaction. The exact physical mechanism for NBTI damage is not clear but often hypothesized to be relevant to the dissociation of  $Si-H$  bonds at the interface and subsequent diffusion of hydrogen in oxide [98]. The up-to-date concept and a reaction-diffusion process ( $R - D$  model [99]) based physical model [100, 101] are discussed in this chapter, upon which a set of new NBTI lifetime model and circuit model are developed.

## 5.2 Accelerated Lifetime Model

From the perspective of process technology, gate oxides are much thinner than before in deep submicron generations and experience increased oxide electric field which is one of the major incentives for NBTI effects. Nitrogen is commonly introduced in pMOSFET's oxide to prevent boron diffusion, increase dielectric constant, suppress gate leakage current and improve hot carrier immunity. However, the inclusion of nitrogen in processes exacerbates NBTI effects [98, 102].

From the perspective of device physics, NBTI becomes a more important reliability concern as device feature sizes shrink below  $0.13\mu m$ . Interface traps and oxide traps generated from the dissociation of interface  $Si-H$  bonds increase carrier surface-related scattering and disturb local electric field in oxide, leading to channel mobility degradation and threshold voltage shift. The electrical effects of NBTI influence on pMOSFETs manifest in decreasing drain saturation current  $I_{dsat}$  and transconductance  $g_m$ , increasing threshold voltage  $V_t$ , and temporarily decreasing off-state current [98, 103].

From the perspective of circuit operation, NBTI is different from HCI in that HCI stresses devices only during the dynamic switching periods when current flows through the device, whereas NBTI stresses devices even when they are in static state operation [104, 105]. The different stress time windows of HCI and NBTI in inverter VTC plot and input-output waveform plot are illustrated in Fig.5.1, which shows that the pMOSFET suffers from NBTI stress when the inverter input voltage is low and output voltage is high, in contrast, the pMOSFET only experiences HCI

stress during the inverter output pulling-up period when  $C_o$  is charging up, while the nMOSFET suffers from HCI stress during the opposite dynamic stage when the inverter output is discharged to low voltage level [106]. The fact that NBTI has a much larger stress time window which even extends to device steady state operation periods leads to the obvious result that duty cycle has much more severe effects on NBTI mechanism, this complicates circuit NBTI behaviors and compels designers to address NBTI effects before circuit fabrication stages.

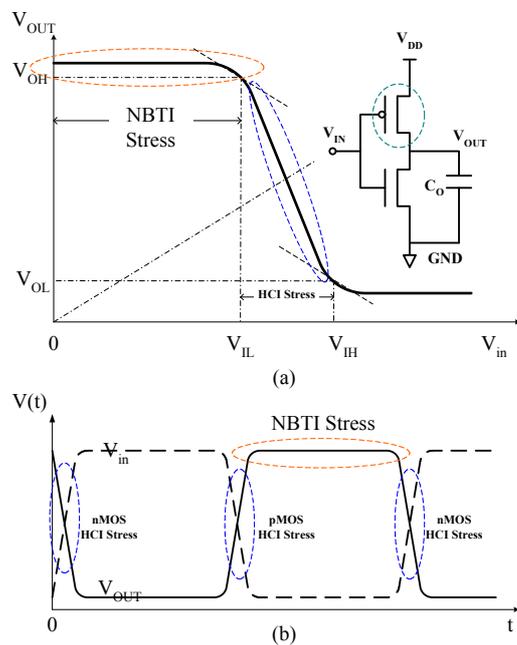


Figure 5.1: The different stress time windows of HCI and NBTI for an inverter in (a) VTC plot and (b) input-output waveform plot. HCI stresses devices only during the dynamic switching periods when both gate voltage and drain voltage are high enough and there is current flowing through the device. NBTI stresses pMOS devices mainly during the period when they are in one of the two static operation states when gate voltage is negative with respect to drain and source voltages.

The most obvious NBTI-induced device degradation phenomenon is the threshold voltage shift  $\Delta V_t(t)$ , therefore, in developing NBTI lifetime models,  $\Delta V_t(t)$  is unanimously used as NBTI degradation monitor to characterize device wearout degree, and accordingly, time to a fixed  $\Delta V_t(t)$  value (e.g.  $50mV$  or  $100mV$ ) is often defined as the NBTI lifetime. As a result of electrochemical reaction-diffusion processes in NBTI, the time dependence of  $\Delta V_t(t)$  is both mathematically derived and experimentally observed to follow a fractional power law relation  $\Delta V_t(t) \propto t^n$ , where the exponent  $n$  ranges from  $0.15 \sim 0.3$  with the typical value of  $0.25$  [104, 107]. The fractional value of  $n$  predicts a saturation behavior at long time  $t$  which is observed in most NBTI experimental work. The voltage dependence of  $\Delta V_t(t)$  is phenomenologically modeled with an exponential law  $\Delta V_t(t) \propto \exp(\beta V_G)$  [108, 109, 110]. The temperature dependence of  $\Delta V_t(t)$  is empirically modeled with the well-know Arrhenius law  $\Delta V_t(t) \propto \exp(-E_a/\kappa T)$  [109, 110]. Taking into account all the above relations,  $\Delta V_t(t)$  is often modeled as [107, 111]:

$$\Delta V_t(t) \propto \exp(\beta V_G) \exp(-E_a/\kappa T) t^n \quad (5.1)$$

NBTI lifetime  $t_f$  is defined as the time to a fixed  $\Delta V_t(t)$  value, therefore, by rearranging equ.(5.1) we obtain a frequently used NBTI lifetime model:

$$t_f = A_0 \exp(-\beta' V_G) \exp(E'_a/\kappa T) \quad (5.2)$$

where  $\beta' = \beta/n$ ,  $E'_a = E_a/n$ , and  $A_0$  is a process constant.

In deriving equ.(5.2), the assumption of exponential law for voltage dependence is not justified and it does not fit recent experimental data very well. Therefore, it is necessary to develop a more suitable acceleration law for NBTI voltage dependency.

A power law voltage acceleration model is developed in the previous work [112]. A phenomenological DC model suggests that shifts in threshold voltage result from the increase in positive fixed charge  $\Delta N_f(t)$  and the generation of donor type interface traps  $\Delta N_{it}(t)$  in the lower half of silicon bandgap [113]:

$$\Delta V_t(t) \propto \frac{q}{C_{ox}}(\Delta N_f(t) + \Delta N_{it}(t)) \quad (5.3)$$

where  $C_{ox}$  is the oxide capacitance. For ultrathin oxide,  $\Delta N_f(t)$  and  $\Delta N_{it}(t)$  are determined by temperature  $T$ , oxide electric field  $E_{ox}$ , oxide thickness  $t_{ox}$  and stress time  $t$ :

$$\Delta N_{it}(t) \propto E_{ox}^m t^{n_1} \frac{1}{t_{ox}} \exp\left(-\frac{E_{a_1}}{\kappa T}\right) \quad (5.4)$$

and

$$\Delta N_f(t) \propto E_{ox}^m t^{n_2} \exp\left(-\frac{E_{a_2}}{\kappa T}\right) \quad (5.5)$$

where  $n_1 = 0.25$ ,  $E_{a_1} = 0.2eV$  for  $\Delta N_{it}(t)$ , and  $n_2 = 0.14$ ,  $E_{a_2} = 0.15eV$  for  $\Delta N_f(t)$ , respectively, and  $m = 1.5$  for both cases [98]. Equ.(5.4) shows thickness dependence of  $\Delta N_{it}(t)$  on  $t_{ox}$ , while equ.(5.5) means  $\Delta N_f(t)$  is thickness independent. These dependencies prompt an assumption that for smaller  $t_{ox}$ ,  $\Delta N_{it}(t)$  will dominate over  $\Delta N_f(t)$  in equ.(5.3) (this assumption is supported in [114]). Substituting equ.(5.4) into equ.(5.3) and neglecting  $\Delta N_f(t)$ ,  $C_{ox}$  and  $t_{ox}$  will cancel each other (because  $C_{ox} = \varepsilon_{ox}/t_{ox}$ ) in equ.(5.3). If we replace the oxide electric field  $E_{ox}$  with the gate bias voltage  $V_{gs}$  (for  $p^+$  poly-Si gate pMOSFETs,  $E_{ox} = (V_{gs} - 0.2V)/t_{ox} \approx V_{gs}/t_{ox}$  according to equ. (22) in [98]), then we get a new NBTI lifetime model:

$$t_f = A_1 \left(\frac{1}{V_{gs}}\right)^\gamma \exp\left(\frac{E_a}{\kappa T}\right) \quad (5.6)$$

where  $E_a$  is activation energy,  $A_1$  is process related constant,  $\gamma$  is voltage acceleration factor. This voltage power law relation is also reported in [115]. In literature, the typical value of  $E_a$  is reported as  $0.9 \sim 1.2eV$ , and the  $\gamma$  value is about  $6 \sim 8$  [115, 116].

Quick development of NBTI testing and analyzing techniques have discovered some new phenomena of NBTI effects including dynamic recovery effect [107, 108, 109] and  $\Delta V_t(t)$  saturation effect [109]. These new phenomena require new physics-based lifetime models to account for and predict NBTI impact on circuit performance and functionality. Based on the model proposed by Zafar [100, 101], a new NBTI lifetime model is developed by taking into account degradation physics and statistical mechanics. This new model provides new statistical explanation for  $\Delta V_t(t)$  saturation effect and physical explanation for dynamic recovery effect in the same framework. Based on the same Zafar model, a new NBTI circuit model is constructed which is the first circuit level model in this area for modeling NBTI effect on circuit functionality.

According to [101], by applying statistical mechanics to calculating the decrease in interfacial  $Si - H$  density as a function of stress conditions, we can mathematically derive a new time dependence of  $\Delta V_t(t)$  as:

$$\Delta V_t(t) = \Delta V_{max} [1 - e^{-(\frac{t}{\tau})^\beta}] \quad (5.7)$$

where  $\Delta V_{max}$ ,  $\tau$  and  $\beta$  are three model parameters. The parameter  $\Delta V_{max}$  is the maximum  $\Delta V_t(t)$  shift that would occur when all the interfacial  $Si - H$  bonds have been depassivated. The parameter  $\tau$  is the time when  $\Delta V_t(t)$  increases to 63% of

$\Delta V_{max}$  and therefore is a measure of the NBTI degradation rate. The parameter  $\beta$  ( $0 < \beta < 1$ ) is a measure of dispersion in hydrogen diffusion and its value decreases from 1 to 0 as dispersion increases.  $\beta$  is independent of stress oxide field  $E_{ox}$  [101].

$\tau$  and  $\Delta V_{max}$  have been derived in [101] as:

$$\tau = B_1 E_{ox}^{-\frac{1}{\beta}} \quad (5.8)$$

and

$$\Delta V_{max} = B_2 \left[ \frac{1}{1 + 2 \exp(-\frac{E_1}{\kappa T})} + \frac{1}{1 + 2 \exp(-\frac{E_2}{\kappa T})} \right] \quad (5.9)$$

where  $B_1$  and  $B_2$  are model prefactors.  $E_1$  and  $E_2$  are material and oxide electric field dependent parameter. Their values are given as:

$$E_1 = E_{it} - E_g + E_F \quad (5.10)$$

and

$$E_2 = E_{fx} - E_F + \gamma E_{ox}^{\frac{2}{3}} \quad (5.11)$$

where  $E_{it}$  and  $E_{fx}$  are trap energy level at the oxide/*Si* interface and trap energy in the oxide, respectively,  $E_F$  is Fermi energy with respect to valence band edge in bulk *Si*,  $E_{ox}$  is the applied electric field across the oxide,  $\gamma$  is a constant and  $\gamma E_{ox}^{2/3}$  represents the decrease in the electronic energy due to band bending in the substrate. A set of typical values for these parameters are given in [101]:  $E_{it} = 0.24eV$ ,  $E_{fx} = -0.16eV$ ,  $E_g = 1.12eV$ ,  $E_F = 0.98eV$ ,  $\gamma = 6.64 \times 10^{-7}$ . Based on these values, we obtain  $E_1 = 0.10eV$ , and  $E_2 = 0.14eV$  (if assume  $V_{ox} = 1V$  and  $t_{ox} = 10nm$ ).  $E_1$  is a process determined parameter, while  $E_2$  is a circuit operation dependent parameter due to the fact that  $V_{ox}$  is a function of  $V_{gs}$ .

If we define  $F(t) = \Delta V_t(t)/\Delta V_{max}$ , then we can rewrite equ.(5.7) in the form:

$$F(t) = 1 - e^{-(\frac{t}{\tau})^\beta} \quad (5.12)$$

Equ.(5.12) is exactly the same as Weibull function. If we define  $f(t) = \partial F(t)/\partial t$ , then  $f(t)$  represents the rate-of-change in  $\Delta V_t(t)$  (normalized to  $\Delta V_{max}$ ). Based on the above transformations, one can explain NBTI time dependent degradation behaviors (power law at initial period followed by a gradual saturation effect) with Weibull statistics. When  $(t/\tau)^\beta$  is very small (corresponds to initial NBTI stress), with the mathematical approximation  $e^{-x} \approx 1 - x$ , equ.(5.12) can be simplified to:

$$F(t) = 1 - e^{-(\frac{t}{\tau})^\beta} \approx [1 - (1 - (\frac{t}{\tau})^\beta)] = (\frac{t}{\tau})^\beta \propto t^\beta \quad (5.13)$$

Equ.(5.13) shows that at the initial state, NBTI-induced  $\Delta V_t(t)$  follows a power law time dependency.

According to Weibull statistics, if the slope parameter  $\beta$  is smaller than 1, the probability density function  $f(t)$  will decrease with time  $t$ . In equ.(5.12),  $\beta$  is always smaller than 1 (i.e.  $0 < \beta < 1$ ), which means that the rate-of-change in  $\Delta V_t(t)$  (normalized to  $\Delta V_{max}$ ) will decrease with time. Therefore, at the very long time  $t$ ,  $\Delta V_t(t)$  will gradually saturate. The above Weibull equivalent explanations justifies the validity of equ.(5.7) from statistics point of view. From equ.(5.7), we can derive a new NBTI lifetime model which perfectly explains NBTI dynamic recovery effects.

Rearranging equ.(5.7) and solving for time  $t$ , we obtain:

$$t = \tau \left[ \ln \frac{1}{1 - \frac{\Delta V_t(t)}{\Delta V_{max}}} \right]^{\frac{1}{\beta}} \quad (5.14)$$

By substituting equ.(5.8) into equ.(5.14), we rewrite equ.(5.14) as:

$$t = B_1 E_{ox}^{-\frac{1}{\beta}} \left[ \ln \frac{1}{1 - \frac{\Delta V_t(t)}{\Delta V_{max}}} \right]^{\frac{1}{\beta}} \quad (5.15)$$

The relations between  $E_{ox}$  and gate voltage  $V_{gs}$  is given as (according to equ. (21) in [98]):

$$E_{ox} = \frac{V_{gs} - V_{FB} - \phi_s}{t_{ox}} \approx \frac{V_{gs} - 0.2V}{t_{ox}} \quad (5.16)$$

where  $V_{FB}$  is flat-band voltage and  $\phi_s$  is surface potential.

Equ.(5.16) can be written in a general form as:

$$E_{ox} \propto V_{gs} - \alpha \quad (5.17)$$

where  $\alpha$  is a technology related potential constant with typical value of 0.2V for pMOSFETs with  $p^+$  poly-gate.

Equ.(5.15) can be transformed to equ.(5.18) by substituting with equ.(5.17):

$$t = B_1 (V_{gs} - \alpha)^{-\frac{1}{\beta}} \left[ \ln \frac{1}{1 - \frac{\Delta V_t(t)}{\Delta V_{max}}} \right]^{\frac{1}{\beta}} \quad (5.18)$$

According to the mathematical approximation that if  $x$  is very small, then  $\ln[1/(1 - x)] \approx x$  (e.g. if  $x = 0.1$ ,  $\ln[1/(1 - x)] = 0.1054$ , the relative error is only 5.4%), we can further simplify equ.(5.18). Because most device service times at normal use conditions are much shorter than device's end-of-life lifetimes, it is reasonable to assume  $\Delta V_t(t)/\Delta V_{max}$  to be a very small quantity (the  $1/\beta$  exponent of it tends to further shrink the difference between  $\ln[1/(1 - x)]$  and  $x$ ). Therefore, equ.(5.18) is reduced to:

$$t = B_1 (V_{gs} - \alpha)^{-\frac{1}{\beta}} \left[ \frac{\Delta V_t(t)}{\Delta V_{max}} \right]^{\frac{1}{\beta}} \quad (5.19)$$

Substituting equ.(5.9) into equ.(5.19) and neglecting the effect of  $\alpha$  on  $V_{gs}$  shift (suppose  $V_{gs}$  is much larger than  $0.2V$ ), we obtain a new physics and statistics based NBTI lifetime model:

$$t_f = A_{NBTI} V_{gs}^{-\frac{1}{\beta}} \left[ \frac{1}{1 + 2 \exp(-\frac{E_1}{\kappa T})} + \frac{1}{1 + 2 \exp(-\frac{E_2}{\kappa T})} \right]^{-\frac{1}{\beta}} \quad (5.20)$$

where the typical value of  $\beta$  is 0.3 [101],  $E_1$  and  $E_2$  are given by equ.(5.10) and equ.(5.11), respectively.

Equ.(5.20) is the mathematical transformation of equ.(5.7), therefore, it inherits all the merits of equ.(5.7). This means our new NBTI lifetime model inherently accounts for NBTI  $\Delta V_t(t)$  power law and saturation behaviors having been discussed before. Another main feature of this new model is its accountability for NBTI dynamic recovery and AC effects. Traditional NBTI analysis neglects these important new effects obviously observed from latest experimental work which lead to relaxed NBTI degradation [108, 117]. If these effects are not considered, an over pessimistic NBTI lifetime will be extrapolated, which exacerbates the already over-depredated reliability margins. In dynamic digital circuit operations or analog circuit AC operations, NBTI effect can be roughly treated as a two-step stress process: a high stress period and a low stress recovery period. According to this new NBTI lifetime model,  $E_2$  is voltage dependent (from equ.(5.11)), therefore,  $E_2$  will be larger at high stress period and smaller at low stress period. According to equ.(5.20), a higher  $E_2$  leads to a shorter  $t_f$  and a lower  $E_2$  leads to a longer  $t_f$ . The final  $t_f$  for the whole process is the interposition of these two processes. Therefore, equ.(5.20) both reveals the origins and provides a prediction method for NBTI dynamic recovery and AC

effects. The above discussion proves that the new NBTI lifetime model outperforms other peer models (equ.(5.2), equ.(5.6) and the model in [118]) in that it accounts for nearly all known NBTI effects in a unified framework for reliability analysis.

### 5.3 Failure Equivalent Circuit Model

To date, the dominating work on NBTI has been concentrated on discrete transistor parameter drift, rather than on circuit performance degradation [105, 119]. Recently, the interest of NBTI community has been gradually elevated to characterizing impacts of NBTI on digital circuit reliability [104, 105, 119, 120, 121, 122], and on analog and RF circuit reliability [123, 124, 125]. Reddy et al [105, 119] developed an NBTI circuit degradation model to investigate the first-order impact of NBTI-induced pMOSFET degradation on ring oscillator and SRAM circuit performances. This model establishes a simple relationship between inverter propagation delay and device threshold voltage shift, thereby enabling circuit frequency degradation simulation due to NBTI-induced device parameter drift. Compared to HCI reliability, it is more difficult to identify NBTI critical subcircuits because of the obvious absence of effective NBTI circuit model. Furthermore, NBTI degrades device parameters even when they are in static state, therefore, NBTI critical subcircuits must be identified as early as possible in the design cycle [120]. This is supported by a simple example: as we know, DC biased circuits are very important for circuit operation, especially for analog and mixed-signal circuits, but they are prone to NBTI degradation, so if the most NBTI-sensitive subcircuits in biasing networks were not

identified and properly designed, the overall circuit could not be NBTI-robust.

It is very important to be able to simulate the impact of NBTI at circuit level using SPICE simulation [120]. In most of the work on NBTI SPICE simulation, it is performed in such a way that degraded circuit behaviors are simulated with SPICE transistor model parameter  $V_t$  being arbitrarily perturbed and shifted by a fixed value [105]. This kind of simulation method cannot physically relate circuit performance degradation to device NBTI wearout process in dynamic operation situations because the parameter  $t$  (NBTI stress time) is not set in. The most effective way to build up this kind of relation is through NBTI circuit model. However, to our best knowledge, there is no electrical model of this kind existed in literature. Based on the previously introduced Weibull law time dependent  $\Delta V_t(t)$  model (equ.(5.7)), a new NBTI circuit model is proposed which is the first electrical model relating the time dependent NBTI physical degradation parameter  $\Delta V_t(t)$  to lumped electrical model elements, thereby enabling effective and quick NBTI circuit reliability simulation.

As mentioned before, the most severe NBTI effect is pMOSFET threshold voltage increase  $\Delta V_t(t)$ , which is equivalent to pMOSFET absolute gate-to-source voltage decrease. Therefore, if splitting the pMOSFET gate connection and adding a gate resistance  $R_G$  between the original gate biasing point  $G$  (voltage at this point is preserved as before by biasing circuit) and the pMOSFET immediate gate terminal  $G'$ , and constructing a gate leakage current flowing mechanism (voltage controlled current sources between gate-to-drain and gate-to-source), then the gate leakage current will flow through this gate resistance  $R_G$  and increase the pMOSFET

effective gate voltage at point  $G'$ . Because pMOSFET source is held at the fixed highest potential, the inclusion of  $R_G$  and gate leakage current leads to the decrease of pMOSFET absolute gate-to-source voltage, thereby imitating the NBTI threshold voltage degradation. Based on this concept, the NBTI circuit model is constructed and shown in Fig.5.2.

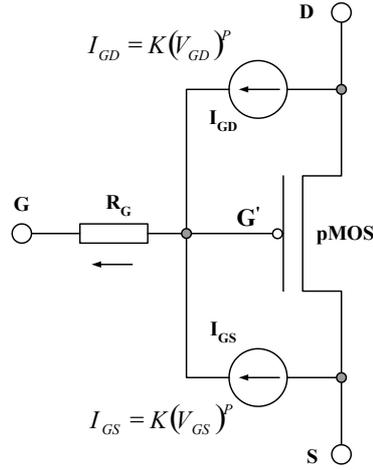


Figure 5.2: MaCRO NBTI circuit model. NBTI-induced pMOSFET threshold voltage increase is modeled as absolute gate-to-source voltage decrease. Gate tunneling current flowing through the gate resistance  $R_G$  leads to the increase of voltage at point  $G'$ . This corresponds to the decrease of pMOSFET absolute gate-to-source voltage and therefore mimics the threshold voltage degradation effect. Gate tunneling current is modeled with two voltage controlled current sources which follow the form of a power law relation as:  $I = KV^P$ .

In this model,  $R_G$  is a voltage dependent resistance because gate leakage currents are voltage dependent.  $R_G$  is also a time dependent resistance because voltage

drop across  $R_G$  at any specific time  $t$  is equal to threshold voltage shift  $\Delta V_t(t)$  which is time dependent. According to [126], gate leakage current due to oxide breakdown conduction can be modeled as gate-to-diffusion leakage with a power law dependence of the formula  $I = KV^p$  (where  $K$  and  $p$  are fitting parameters). The same power law voltage dependency, shown in Fig.5.2, is adopted in modeling gate leakage currents. As a result, for the gate-to-drain leakage current,  $I_{GD} = K(V_{GD})^p$ , and for the gate-to-source leakage current,  $I_{GS} = K(V_{GS})^p$ . In MaCRO, the default value of  $p$  is set to 5, and the default value of  $K$  is  $3 \times 10^{-6}$  [126].

In Fig.5.2, the voltage drop across  $R_G$  is:

$$V_{R_G}(t) = V_{G'} - V_G = \Delta V_G(t) = (I_{GD} + I_{GS})R_G \quad (5.21)$$

Threshold voltage degradation  $\Delta V_t(t)$  due to NBTI is already given by equ.(5.7), therefore, from the relation  $\Delta V_G(t) = \Delta V_t(t)$ , we can obtain an analytical solution for  $R_G$ :

$$R_G = \frac{\Delta V_{max}}{KV_{GD}^p + KV_{GS}^p} [1 - e^{-(\frac{t}{\tau})^\beta}] \quad (5.22)$$

The typical values and extraction methods for the model parameters  $\Delta V_{max}$ ,  $K$ ,  $p$ ,  $\tau$  and  $\beta$  have been given and discussed during the process of deriving equ.(5.22).

One of the most important points shown in Fig.5.2 is that this new model is much better than a simple model which only inserts a voltage source between  $G$  and  $G'$  representing threshold voltage shift in that it inherently incorporates both NBTI and possible oxide breakdown effects.

For nMOS Positive Bias Temperature Instability (PBTI) circuit model, a similar structure as that for pMOS NBTI shown in Fig.5.2 can be constructed except

that all current flowing directions are reversed, and the model fitting parameters of the threshold voltage model  $\Delta V_t$  (equ.(5.7)) are determined from nMOS PBTI stress testing. For the two current sources ( $I_{GD}$  and  $I_{GS}$ ) in nMOS PBTI circuit model, a better gate leakage model proposed by Lee et al [127] is adopted:

$$I_{GS} = \frac{1}{2}AL \exp(\alpha V_{GS} - \beta t_{ox}^{-\gamma}) \quad (5.23)$$

and

$$I_{GD} = \frac{1}{2}AL \exp(\alpha V_{GD} - \beta t_{ox}^{-\gamma}) \quad (5.24)$$

where  $I_{GS}$  and  $I_{GD}$  are in unit  $\mu A$ ,  $L$  is effective channel length in nanometer,  $t_{ox}$  is oxide thickness in nanometer,  $A = 127.04$ ,  $\alpha = 5.61$ ,  $\beta = 10.6$  and  $\gamma = 2.5$ . These typical values for nMOSFETs were obtained by fitting industrial data and found to good for technologies across many generations up to  $0.13\mu m$ . These new leakage models are able to maintain good stability in SPICE simulation [127].

## 5.4 Implementation in MaCRO

In MaCRO, when applying the NBTI circuit model to SPICE circuit reliability simulation, designers first perform SPICE simulation without considering NBTI damage. SPICE will predict average values of  $V_{GD}$  and  $V_{GS}$ , and average values of  $I_{GD}$  and  $I_{GS}$ . Then from equ.(5.22), they can determine the gate resistance  $R_G$  at any specified time  $t_s$ . With this  $R_G$  value and the voltage dependent current sources  $I_{GD}$  and  $I_{GS}$ , they substitute the most NBTI damaged pMOSFETs (identified through the same procedure as those of HCI and TDDDB) with the NBTI circuit

model, then perform another round SPICE simulation. The circuit performance and functionality at the time  $t_s$  can be analyzed with this NBTI circuit model. For PBTI damaged nMOS devices, designers can use similar models with proper voltage/current polarities and device parameters, especially, the sign of the first activation energy term in lifetime equ.(5.20) should be negative to account for charge subtraction relation between interface trapped charge and oxide trapped charge. For simplicity, PBTI on nMOSFET is often neglected in circuit reliability simulation due to its very weak influence.

Another advantageous feature of the NBTI circuit model is its expandability. When MOSFETs scale into ultrathin oxide regimes (i.e.  $t_{ox} < 5nm$ ), electron direct tunneling mechanism will dominate gate leakage generation. As a result, the above voltage power law dependence of leakage current may be not valid any more. In [128], gate leakage due to direct tunneling through ultrathin oxide ( $t_{ox} < 2nm$ ) is characterized with an explicit surface potential model with quantum-mechanical corrections, and a compact gate leakage current model feasible for SPICE simulation is developed. New  $I_{GD}$  and  $I_{GS}$  leakage models in terms of surface potential are also proposed. Based on these new leakage current models, with minor modifications, MaCRO can be easily expanded to model NBTI effects in future technologies.

For both NBTI circuit model (Fig.5.2) and PBTI circuit model, in real circuit operation situations, the flowing of the current source  $I_{GD}$  may reverse if gate voltage crosses over drain voltage. This current source reverse effect is not a problem for the TDDB circuit model, because TDDB creates a physical path between gate and drain, and the gate-to-drain current can flow in either direction depending on

relative magnitude of gate and drain voltages. Whereas in the NBTI circuit model, this current path is visualized and not physically exists, so the current source reverse problem needs to be addressed in MaCRO NBTI/PBTI reliability simulation. As discussed before, NBTI and PBTI mainly stress devices during steady state operation periods, so in device dynamic switching periods which are normally short but may lead to current source reverse phenomenon, one can treat the device as no NBTI/PBTI effects at the drain end during these transition periods. In another word, he can disable the current source  $I_{GD}$  when gate and drain voltages cross over in dynamic periods. This countermeasure against current source reverse effect can be easily implemented with SPICE structure control commands (e.g. if-then-else control flow).

The Matlab program for calculating NBTI circuit model parameters in terms of  $0.25\mu m$  technology parameters is included in Appendix A.3.

## 5.5 Conclusion

In this chapter, a new NBTI lifetime model is developed based on an existing physics and statistics based model. This new lifetime model accounts for most experimental observations on NBTI-induced threshold voltage degradation behaviors including fractional power law dependence, saturation phenomenon and dynamic recovery effect. Weibull statistics is used in explaining this NBTI lifetime model, which provides a new thoughtway in understanding NBTI degradation behaviors. A new NBTI circuit model is developed based on physics-of-failure concept. This

NBTI circuit model features simplicity and expandability, and is presumed to be the first NBTI damaged circuit model of this kind in literature.

## Chapter 6

### Electromigration Lifetime Models and Parameter Extraction

#### 6.1 Introduction

Besides the three MOSFET-related wearout mechanisms having been discussed in the previous chapters, Electromigration (EM) is another important failure mechanism inherent to Silicon chips. Keeping pace with the shrinking of MOSFET physical dimensions, both interconnecting layers and geometries of on-chip metalization scale very quickly. This leads to higher current density flowing through the interconnects, exacerbating EM wearout effects on circuit performance and reliability. As a result, even though some new materials with better immunity to EM failures have been used as on-chip interconnects to replace Aluminum (Al), EM is still a major reliability concern, and designers need accurate EM lifetime models to correctly predict device failure rate and derate circuit for long life applications.

However, accurate lifetime model for EM wearout mechanism is not enough, practical algorithms of model parameter extraction are also very important. The pressure to deliver designs to market quickly and reliably fosters the development and application of accelerated tests in electronic product design and reliability qualification. Accelerated test not only serves as the most effective means in developing

and validating lifetime models of wearout mechanisms, but also provides possibility to extrapolate model parameters in a relatively short time. In this chapter, parameter extraction process for EM lifetime models with accelerated test method is discussed. This process can be easily extended and applied to other wearout mechanisms.

## 6.2 Electromigration Failure Physics

EM is the mass transport of a metal wire due to the momentum exchange between the conducting electrons which move in the applied electric field, and the metal atoms which make up of the interconnecting material. EM exists wherever electric current flows through metal wires. With the advent of deep submicron CMOS technologies, on-chip interconnects are stressed with increased current densities. In this case, EM will lead to much shorter times to electrical failure of the interconnects, thereby reducing circuit reliability to an unacceptable level.

EM failure kinetics for different metal structures such as long lines, vias and contacts are different due to their different line widths and material characteristics. Therefore, the kinetics of EM failures for each of these structures must be analyzed separately and evaluated accordingly. The main driving forces for EM failures are current density and temperature, but their acceleration effects follow distinct trends for different EM failure kinetics and depend on whether the failure is nucleation-dominated or growth-dominated. The difference of these two kinds of EM failure kinetics is addressed in the section.

### 6.2.1 Nucleation-Dominated Failure Physics

The difference between nucleation-dominated and growth-dominated failures results from the process used to deposit the metal and the overlying dielectric. Nucleation-dominated failure is typical for structures that do not contain a redundant shunt layer of refractory materials. For example, the failure of an Al-alloy stripe which is terminated by bonding pads and has no barrier metallization is dominated by nucleation. Void nucleation occurs if significant mass transport takes place and sufficient stress is generated. When stress accumulates to a critical level, a void will come into being to reduce the stress in the materials. After the void forms, if there is no shunt layers, an open circuit failure will develop very quickly due to the abrupt release of strain energy. This is hard failure type and can be easily detected. Besides void nucleation phenomenon, there are two other nucleation-dominated failure mechanisms: the stress buildup following Cu depletion in Al/Cu alloys, and passivation cracking induced by compressive stresses which produce extrusions.

### 6.2.2 Growth-Dominated Failure Physics

In contrast, when a void exists in the primary metal conductor, if there is a redundant shunt layer, the initial rapid growth of the void will not produce an open circuit failure because the refractory material can conduct electricity. This structure can withstand very high current densities and temperatures for very long time. This is a soft failure phenomenon and called growth-dominated EM failure. For this kind of soft failure, the 10% shift in resistance of global wiring is normally

chosen as a failure criterion. For a given metal structure, nucleation-dominated failure will happen much sooner than growth-dominated failure, since the damage needs to nucleate before it can grow [129].

### 6.3 Electromigration Lifetime Models

A practical EM lifetime model must realize two important functions. Firstly, it must identify critical stress parameters, provide guidelines to perform accelerated tests, and account for the relations between test results and actual use conditions. Secondly, EM failure behaviors and physics of different metal structures must be taken into account, so that the test results can be extended to real and complex circuits, enabling proper estimation of product reliability. Traditionally, these two functions have been treated separately, but new experimental and research work has led to a general model unifying these two aspects into one framework.

The original Black model is the first accepted EM lifetime model. It is an empirical model for grain-boundary controlled EM failures and fits field data very well. However, the activation energy extracted from experimental data may be inaccurate because this model has not been physically justified [131] and only applies to thin conductor films, whose line width is many times larger than the average grain size [130]. To overcome this activation energy inaccuracy problem, a generalized Black model has been proposed to characterize EM failure behaviors [131]:

$$t_f = A_{EM} J^{-n} T^{-m} \exp\left(\frac{E_{aEM}}{\kappa T}\right) \quad (6.1)$$

where  $A_{EM}$  is process and material related constant,  $J$  is the average current den-

sity,  $\kappa$  is the Boltzmann's constant,  $T$  is temperature in Kelvin, and  $E_{aEM}$  is an experimentally determined activation energy.

The various combinations of  $n$  and  $m$  values are determined by the particular failure physics of an interconnect and its geometry. If considering failure physics: for nucleation-dominated failures,  $n = 2$  and  $m = 0$ , while for growth-dominated failures,  $n = 1$  and  $m = 0$ . If considering interconnect's geometry: for wide lines whose average grain size is much smaller than their line widths,  $n = 2$  and  $m = 0$ , however for narrow lines,  $n = 1$  and  $m = 0$ . When  $n = 2$  and  $m = 0$ , it is the same as the original Black model.

In engineering applications, it is proved that there is no significant difference for which combination of  $n$  and  $m$  values is used, however, calculations show that the case  $n = 2$  and  $m = 2$ , i.e. Shatzkes and Lloyd model (S-L model), produces very good lifetime results with the extrapolated activation energies being reasonably accurate [131]. Nevertheless, there is a drawback in using this S-L model that the inclusion of non-zero  $m$  parameter leads to nonlinear relationship between lifetime and temperature after logarithmic transformation of equ.(6.1), which complicates model parameter extraction process.

## 6.4 Model Parameter Extraction

The lifetime models presented above use the average current density for lifetime projection, therefore, they are good for DC current stress conditions. However, in a real circuit, most on-chip metal wires also experience AC current stresses [133].

Extended EM lifetimes have been observed under bidirectional current stresses, and experiments show that lifetime enhancement in pure AC stresses may be two orders of magnitude larger than that in pure DC stresses [132]. Therefore, in addressing EM effects in real applications, the lifetime model must be able to characterize current density under bidirectional stresses.

#### 6.4.1 Current Density

The current density  $J$  in an interconnect segment can be expressed as [134]:

$$J = \frac{C_{int} \times V}{W \times H} \times f \times \gamma \quad (6.2)$$

where  $C_{int}$  is the interconnect capacitance of a specific node in the circuit,  $V$  is the voltage drop across the interconnect segment,  $W$  and  $H$  are interconnect width and thickness and determined by design rules and technology,  $f$  is the current switching frequency, and  $\gamma$  is the probability that the line switches in one clock cycle.

The interconnect capacitance  $C_{int}$  at any node contains three components: overlap capacitance  $C_{over}$ , lateral capacitance  $C_{lat}$  and fringe capacitance  $C_{fr}$ .

The overlap capacitance  $C_{over}$  is due to the overlap between two conductors in different layers. It is modeled as follows by taking into account the overlap area:

$$C_{over} = C_a \times W \times L \quad (6.3)$$

where  $C_a$  is capacitance per unit area,  $W$  and  $L$  are width and length of the overlap area, respectively.

The fringe capacitance  $C_{fr}$  is due to the coupling effect between two conductors

in different layers. It is modeled as:

$$C_{fr} = 2 \times C_l \times L \quad (6.4)$$

where  $C_l$  is the capacitance per unit length of the edge of the top wire,  $L$  is the perimeter of this wire.

The lateral capacitance  $C_{lat}$  is the capacitance between two conductors in the same layer. The continuous shrinking interconnect dimensions and increasing metallization layers in deep submicron technologies render the overlap and fringe capacitances much larger than the lateral capacitance, therefore, the contribution of  $C_{lat}$  can be neglected. With this assumption, the interconnect capacitance  $C_{int}$  is only governed by  $C_{over}$ ,  $C_{fr}$  and wire geometries:

$$C_{int} = (C_a \times W + 2 \times C_l) \times L \quad (6.5)$$

$C_a$  and  $C_l$  are technology and material dependent constants, they can be extrapolated from technology files.  $W$  and  $L$  are design parameters, they can be obtained from the layout design files. With the availability of these parameters,  $C_{int}$  is easy to be predicted with equ.(6.5) [135]. For instance, in a typical  $0.25\mu m$  technology,  $C_a$  of the first layer Al is about  $30aF/\mu m^2$ , and  $C_l$  is about  $40aF/\mu m$ , for an Al wire of  $10cm$  long and  $1\mu m$  wide and routed on the first Al layer,  $C_{int}$  is predicted to be  $11pF$  [136].

In equ.(6.2),  $\gamma$  is the signal activity which associates with the average number of transitions occurring at any particular node of a circuit. It is a measure of the stress that can cause failures in digital circuits. In real applications, how to determine when and how often transitions occur at a node is a difficult work because

signal transitions at circuit internal nodes depend on the kind and sequence of input vectors. In normal operations, both input vectors' kind and sequence vary widely, they may not follow regular patterns. Therefore, an accurate estimation of signal activity has to take into account many factors including signal correlations, simultaneous switching, and probabilistic techniques [137]. Because of this complexity, recently, SPICE simulator starts to be used in EM analyses to “probe” the current waveforms on interconnect segments. While the accuracy of  $\gamma$  from SPICE simulation may be limited by the finite combinations of input stimuli, the deviation is expected to be insignificant.

Until now, all the parameters in equ.(6.2) have been discussed and they can be extracted either from SPICE simulation or from technology files. What follows is the way to determine the model parameters, i.e. current acceleration factor  $n$  and activation energy  $E_{aEM}$ , in equ.(6.1). The detailed process, guidelines and examples on extraction of these parameters have been documented in JESD63 [138], even so, it is necessary to recapitulate the most important aspects in this section and discuss some missing points.

## 6.4.2 Current Acceleration Factor and Activation Energy

### 1) Overview

The EM model parameter extrapolation methods proposed in JESD63 provide procedures to use linear regression analysis in calculating model parameters for thin-film metal interconnects used in modern integrated circuits. These methods

are based on the fact that MTTF data from accelerated tests of sample wires can be satisfactorily approximated by equ.(6.1), assuming  $m = 0$ . The two model parameters need to extrapolate are the exponent  $n$  for current density  $J$ , and the activation energy  $E_{aEM}$  for temperature. The linear regression analysis predicts  $n$  and  $E_{aEM}$  directly from testing data of sample wires by plotting the data in logarithmic scales and measuring the slopes of fitting lines.

The extrapolation process requires existing failure-time data or MTTF data. In testing this lifetime data, sample wires must be tested at no less than three different EM stress levels if varying only the current density or only the temperature. However, if both current density and temperature are varied during the tests, the lifetime data must come from at least four different combinations of these two stress factors and constitute four corners of a quadrangle when plotting the result. This is named Matrix Stressing Method (MSM). If within the temperature range used to test  $E_{aEM}$ , significant migration occurs in the lattice, then the assumption  $m = 0$  is not valid, and a plot of  $\ln MTTF$  vs.  $1/T$  may demonstrate obvious nonlinearity. This usually applies to the metal interconnect lines whose median grain size is comparable with or larger than the line width. For this kind of lines, it is necessary to perform more complicated accelerated tests to extrapolate  $m$  parameter, or evaluate if tests can be conducted at different temperature ranges to separate competing mechanisms.

## 2) Current Acceleration Factor $n$

If during the tests, the current density is varied while the temperature is kept

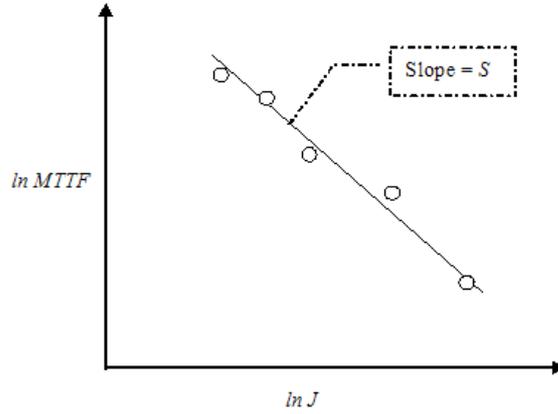


Figure 6.1: Plot of  $\ln MTTF$  vs.  $\ln J$  to illustrate the fitting method in extracting  $n$  and judge the linearity of the dependence [138].

constant, then  $MTTF$  will be proportional to  $J^{-n}$  and hence:

$$\ln MTTF = -n \ln J + B \quad (6.6)$$

or

$$\ln MTTF = S \ln J + B \quad (6.7)$$

where  $B$  is a constant, and  $S$  is the slope of the fitting line. A plot of  $\ln MTTF$  vs.  $\ln J$  will display data points closely aligned along a straight line. This is illustrated in Fig.6.1. A linear regression analysis of the  $\ln MTTF$  vs.  $\ln J$  data pairs will yield a least square fitting line to the data with slope  $S$ . The sample estimate for  $n$  is obtained from the relation  $n = -S$ .

### 3) Activation Energy $E_{aEM}$

If the temperature of the test lines is changed while the current density is kept

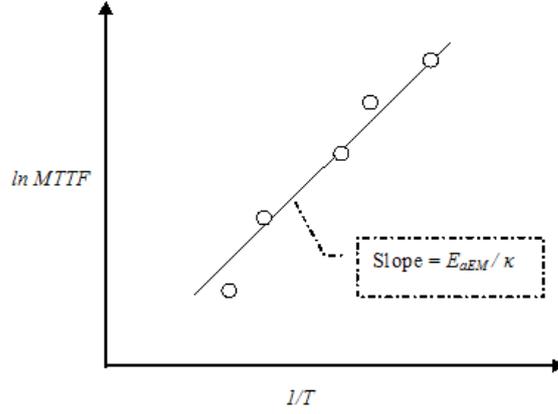


Figure 6.2: Plot of  $\ln MTTF$  vs.  $1/T$  to illustrate the fitting method in extracting  $E_{aEM}$  and judge the linearity of the dependence [138].

constant, then  $MTTF$  will be proportional to  $\exp(E_{aEM}/\kappa T)$  and hence:

$$\ln MTTF = \frac{E_{aEM}}{\kappa} \times \frac{1}{T} + C \quad (6.8)$$

where  $C$  is a constant. A plot of  $\ln MTTF$  vs.  $1/T$  will display data points closely aligned along a straight line. This is illustrated in Fig.6.2. Similarly, a linear regression analysis of the  $\ln MTTF$  vs.  $1/T$  data will yield a least square fitting line to the data as well as the confidence interval for the slope, which is  $E_{aEM}/\kappa$ . The method depends critically on the assumption of linearity, i.e.  $m = 0$ . A very simple way in assessing the validity of this assumption is visual inspection of the plotted data.

### 6.4.3 Temperature Effects

EM is very sensitive to temperature. The maximum current allowed in a thin film conductor is a function of temperature. The higher the temperature, the less

current stress can be tolerated without leading to EM failures. The EM margin in terms of temperature depends on the failure physics, i.e. nucleation-dominated or growth-dominated, and the dominant diffusion mechanism.

In conducting EM stress tests, it is necessary to understand that the real temperature of a sample wire is the sum of (a) the ambient temperature provided by an oven or a hot chuck if for wafer-level tests, (b) the temperature increase due to power dissipation within the sample wire caused by the stress current, and (c) the temperature increase due to power dissipation elsewhere on the wafer or chip, which also elevates the sample wire's temperature. Therefore, during the tests, temperature of sample wires is expected to change with time. Furthermore, resistance of sample wires may fluctuate during the tests, and values of some other wire-associated components can vary with the changing temperature. These variations often lead to the deviation of extracted temperature from its real value. Therefore, in EM lifetime tests, temperature as a stress parameter must be properly controlled and accurately extrapolated.

One simple method to account for the temperature deviations during EM stress tests is adding the observed average temperature rise to the actual oven temperature. The oven temperature is set by measuring the resistance of another resistor (identical to those under stress test in each package) at low current conditions in which Joule heating is negligible [132].

#### 6.4.4 Example of Typical Values

In [133], a piece of on-chip metal wire is tested which connects the two drains in a comparator circuit. The metal wire has geometry of  $1.8 \times 0.85 \times 250mm^3$ . The current flowing through each interconnect in the circuit is estimated from the Cadence-extracted netlists. The calculated current density values are on the order of  $1 \times 10^{10}A/m^2$ . For the specific metal wire being examined, its current density is  $5.2 \times 10^{10}A/m^2$ . The activation energy for Al wires at  $300K$  is extrapolated to be  $0.95eV$  and the current density acceleration factor is selected as  $n = 2$ . With the above information, the MTTF of this metal wire is predicted to be  $1.16 \times 10^{10} \times A$  (in unit Second), where  $A$  is the scaling factor. This simple simulation and calculation procedure can replace the actual yield analysis which requires physical test circuits and chips, thereby saving both development cost and time.

#### 6.5 Electromigration of Copper Wires

Copper (Cu) has lower sheet resistivity and much lower EM failure rate than the traditional metallization material Al, therefore, Cu and low- $K$  intermetal dielectrics begin to replace Al and  $SiO_2$  dielectrics in deep sub-micron CMOS technologies. Although Cu/Low- $K$  materials enable further improvement in circuit speed and EM lifetime, the reliability and yield issues associated with integration of these materials by dual-damascene Cu processing have proven to be more challenging than predicted [139], and the previous EM lifetime parameters for Al material cannot be directly applied to Cu interconnects.

EM effects of on-chip Cu wires plated with dual-damascene processes have been investigated by Hu et al [140]. The widths of these Cu wires vary from  $0.24\mu m$  to  $1.3\mu m$ . Void growth at the cathode and protrusions at the anode of the wires are found to be the main cause of EM failure. The failure lifetime is observed to decrease linearly with reduction of the cross-sectional area of the wires. The factor  $n$  for current density  $J$  in  $0.28\mu m$  wide wires is found to increase from 1 to 2 as  $J$  increases beyond  $25mA/\mu m^2$ . The measured activation energy  $E_{aEM}$  for Cu wires varies widely from  $0.7eV$  to  $1.0eV$  [140].

## Chapter 7

### Derating Voltage and Temperature for Reliability

#### 7.1 Introduction

CMOS technology evolution in the past several decades has been driven by Moore's Law to continuously challenge the scaling limits for higher speed, density and yield. Many distinct scaling theories have been proposed to improve device performance, in which constant-field scaling is the most important one and provides fundamental guidelines to properly scale device physical and geometrical parameters without introducing deleterious high field effects. However, the industry has not followed an exact constant-field scaling because some unshrinkable parameters have prevented the power supply voltage from proportionately scaling with the physical geometries. This leads to the result that the electric fields and current densities in MOSFETs have increased over the generations instead of being maintained constantly. Higher electric fields can cause many reliability problems including HCI, TDDB, Gate-Induced Drain Leakage (GIDL), to name a few [141], while the increased current density will exacerbate EM-related failure mechanisms. As internal stresses of modern semiconductor devices continue to increase, the likelihood of their time dependent wearout and failure also increases. This trend imposes much more

pressure on both designers and manufacturers to deliver qualified products for long life applications.

Technology is mainly driven by a few fast-moving markets such as wireless communication systems and entertainment electronics, in which devices are customized and fabricated to explore their performance to the limits, sometimes by sacrificing reliability. As a result, most Commercial-Off-The-Shelf (COTS) devices currently populated in the market have high performance but short lifetimes, which may limit the lifetimes of the systems in long life applications if these COTS devices are incorporated. The lifetime models developed in the previous chapters prompt a way to address this problem: operating devices at reduced voltage, frequency and temperature than their original ratings (i.e. derating). In CMOS circuits, power dissipation is determined by frequency, voltage and temperature. Reduction in voltage will significantly reduce the power dissipation; similarly, reduction in frequency and temperature will also lead to appropriate reduction in power dissipation. There is a positive relation between the peak power dissipation of CMOS digital circuits and many wearout mechanisms, consequently, even though derating voltage, frequency and temperature does degrade the performance of a device, it also reduces the physical stresses on the device, thereby increasing its expected useful life [142, 143].

This chapter addresses one kind of MaCRO applications, which is derating voltage and temperature for reliability improvement. The lifetime models for HCI, TDDB, NBTI and EM have been presented in the previous chapters. From these lifetime models, if all model parameters are calibrated from testing work, MaCRO can accurately predict device and circuit failure rate, and characterize circuit de-

rating behaviors under different voltage and temperature stresses through SPICE simulation. By introducing a unified derating factor, MaCRO simulation is capable of formulating practical derating design guidelines for improving product lifetime and reliability in long life applications.

## 7.2 Circuit Design and Simulation

A 17-stage ring oscillator consisting of CMOS inverters and interconnecting capacitors is simulated as an example to investigate voltage and temperature derating effects. CMOS ring oscillator has been widely used as test circuit for monitoring process variations and characterizing reliability behaviors because its oscillating frequency is sensitive to SPICE model parameters. Fig.7.1 shows the schematic diagram of the 17-stage ring oscillator. BSIM3v3 model is used to characterize the MOSFETs  $Q_n$  and  $Q_p$ , and the model parameters are taken from TSMC  $0.18\mu m$  CMOS process. TSMC  $0.18\mu m$  CMOS process supports  $1.8V$  and  $3.3V$  applications.  $1.8V$  technology is widely used in general purpose and low power design, while  $3.3V$  is used for high-quality mixed-signal or RF devices. In the following simulation, the rated value of power supply voltage  $V_{DD}$  is chosen as  $3.3V$  in order to set a wider voltage derating range. To obtain symmetrical transfer characteristics, the device gate widths ( $W_n$  of  $Q_n$  and  $W_p$  of  $Q_p$ ) are designed to follow the well-known relationship:

$$\frac{W_p}{W_n} = \frac{I_n}{I_p} = \frac{\mu_n}{\mu_p} \quad (7.1)$$

The extracted values of electron and hole mobilities are  $\mu_n = 263.8cm^2/Vs$

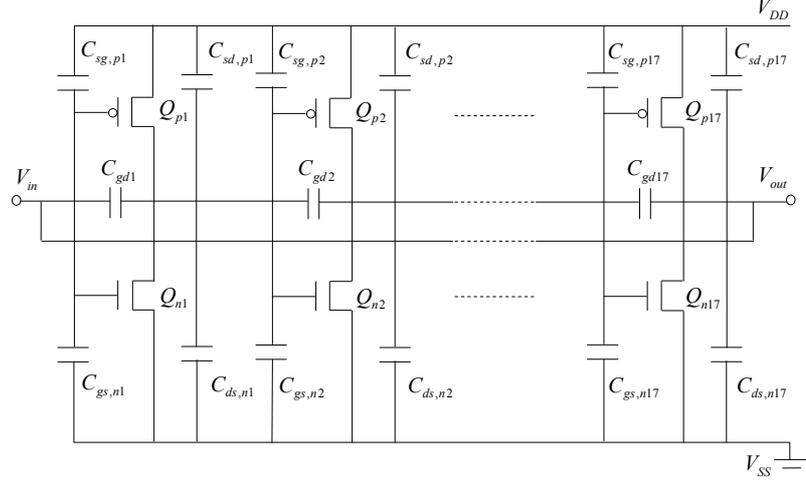


Figure 7.1: The schematic diagram of the ring oscillator which consists of 17-stage CMOS inverters and interconnecting capacitors.

and  $\mu_p = 118.3 \text{ cm}^2/\text{Vs}$ , therefore the gate geometries of  $Q_n$  and  $Q_p$  are designed to be:  $L_n = L_p = 0.35 \mu\text{m}$ ,  $W_n = 1.10 \mu\text{m}$ , and  $W_p = 2.60 \mu\text{m}$ . The overall simulation is divided into three steps to investigate voltage scaling effects, temperature scaling effects and DC transfer characteristics, respectively.

## 7.3 Simulation Results and Analysis

### 7.3.1 Voltage Derating Analysis

The transient analysis is performed by sweeping the power supply voltage  $V_{DD}$  from  $1.0\text{V}$  to  $4.0\text{V}$  with incremental step of  $0.1\text{V}$  to investigate voltage derating behaviors. The ambient temperature is set to  $27^\circ\text{C}$ . When  $V_{DD}$  is scaled, the oscillating frequency monotonically increases from  $80.91\text{MHz}$  to  $418.5\text{MHz}$ .

For a CMOS inverter, if the pull-down delay  $\tau_n$  of nMOSFET is defined as the time for the output voltage decreasing from  $V_{DD}$  to  $V_{DD}/2$ , then  $\tau_n$  can be expressed

as:

$$\tau_n = \frac{CV_{DD}}{2I_{Nsat}} = \frac{CV_{DD}}{\mu_{neff}C_{ox}(W/L)(V_{DD} - V_{tn})^2} \quad (7.2)$$

where  $V_{tn}$  is the threshold voltage,  $\mu_{neff}$  is the electron effective mobility,  $W$  is the channel width and  $L$  is the channel length,  $C$  is the output loading capacitance. When  $V_{DD} \gg V_{tn}$ ,  $\tau_n$  is approximately proportional to the inverse of  $V_{DD}$ . The similar expression can be derived for the pull-up delay  $\tau_p$  of pMOSFET. The transition delay  $\tau$  of the CMOS inverter is the arithmetic average of  $\tau_n$  and  $\tau_p$  (i.e.  $\tau = (\tau_n + \tau_p)/2$ ). Therefore, when  $V_{DD}$  is scaled down in proper range, the operating frequency of the ring oscillator will decrease proportionally. The power consumption of CMOS circuits mainly comes from switching periods in dynamic operation because their static power dissipations are negligible, and the total average power consumption  $P_D$  can be estimated as:

$$P_D = \frac{1}{T} C_L V_{DD}^2 = C_L V_{DD}^2 f \quad (7.3)$$

where  $C_L$  is the total loading capacitance on the chip, and  $f$  is the frequency at which the circuit switches [144].

Fig.7.2 is the simulation results of frequency and power dissipation derating trends with respect to  $V_{DD}$ . Both equ.(7.3) and simulation results show that voltage derating significantly affects power dissipation. When voltage increases 4 times, the frequency increases about 5 times, whereas the power dissipation increases up to 100 times. The net result of the dependence of the power dissipation on the voltage is thus much stronger than a simple quadratic relationship.

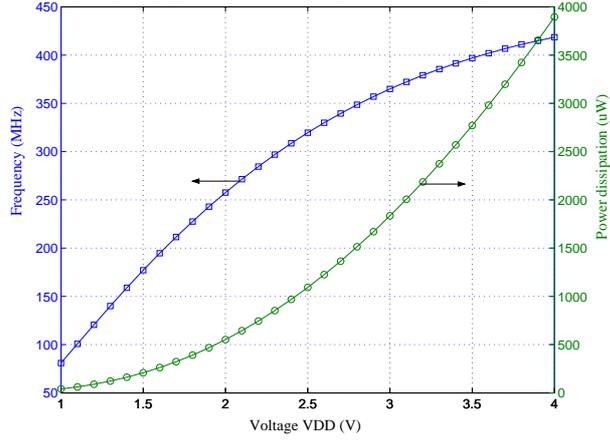


Figure 7.2: The derating relationship of frequency and power dissipation vs.  $V_{DD}$ . When voltage increases 4 times, the frequency increases about 5 times, whereas the power dissipation increases near to 100 times.

The above voltage derating analysis is based on a quasistatic assumption that the device response is quick enough compared with the switching speed of its terminal voltage. This assumption is valid only when the input signal's rise and fall times are much longer than the carrier transit time across the channel. For very short channel devices, the carrier transit time  $\tau_{tran}$  is determined by the carrier saturation transit time  $\tau_{sat} = L/v_{sat}$  or average transit time  $\tau_{avg} = L^2/(\mu_{eff}V_{DD})$ , whichever is larger [141]. For the  $0.18\mu m$  nMOSFET SPICE parameters,  $\tau_{sat} = 3.9ps$ . When  $V_{DD}$  is greater than  $1.2V$ , the electron average transit time across the channel  $\tau_{avg}$  is smaller than  $\tau_{sat}$ , therefore, approximately during the whole range of voltage derating (from  $4.0V$  down to  $1.0V$ ), the device response time, i.e.  $\tau_{tran}$ , is determined by  $\tau_{sat}$  and therefore keeps constant. The simulated minimum switching delay of terminal voltage for CMOS inverter is much larger than  $\tau_{tran}$ . This means the quasi-static assumption is held for the above simulation and the voltage derating behaviors in

light of frequency and power dissipation, given by Fig.7.2, are valid.

The above voltage derating and timing response analyses formulate a guideline for setting proper lower-bounds of  $V_{DD}$  in some special applications. For high frequency, where switching delay is comparable to  $\tau_{tran}$ , and some mixed-signal applications, where long channel devices coexist with short channel devices, if  $V_{DD}$  is derated below some critical value,  $\tau_{tran}$  will be greater than switching delay of device terminal voltage. Thus, the quasi-static assumption would not be valid any more. In these situations, a non-quasistatic model should be incorporated in the simulation to account for possible new voltage derating behaviors.

### 7.3.2 Temperature Derating Analysis

Temperature is another controllable and reliability-sensitive design parameter because a number of important device parameters such as mobility, threshold voltage and saturation velocity are temperature dependent. In order to determine the temperature derating behaviors of frequency and power dissipation, the temperature transient analysis of the same ring oscillator is performed by sweeping the temperature from  $0^{\circ}C$  to  $150^{\circ}C$  with step of  $10^{\circ}C$ .  $V_{DD}$  is set to  $3.3V$  during the process.

Carrier mobility is a well-known temperature dependent parameter. Phonon scattering, surface scattering and impurity scattering are major scattering mechanisms governing the characteristics of carrier mobility and they follow different temperature dependencies. At low temperature, impurity scattering dominates and

the mobility increases with rising temperature, while at high temperature, phonon scattering starts to prevail and the mobility will decrease and follow the trend  $\mu_{eff} \propto T^{-3/2}$ . These competing temperature effects result in a non-monotonic dependence of the mobility on temperature and lead to the existence of a maximum carrier mobility value [145]. According to the discussion in the above section, for long-channel devices operated at very low  $V_{DD}$ , the carrier transit time  $\tau_{tran}$  sets device switching speed and is determined by  $\tau_{avg} = L^2/(\mu_{eff}V_{DD})$ , in which carrier mobility  $\mu_{eff}$  is the only temperature dependent factor. Therefore, derating relation between temperature and frequency of long channel devices operated at low voltage is mainly governed by  $\mu_{eff}$ . Due to the aforementioned non-monotonic dependence of the mobility on temperature, in the derating curve of frequency vs. temperature, there should exist a maximum frequency value at which the mobility is maximal and has relatively weak temperature sensitivity.

If the device channel length is very short and  $V_{DD}$  is high, device operating speed is determined by the interconnecting and parasitic capacitances (refer to equ.(7.2)). In BSIM3v3 model, parasitic capacitances are temperature dependent but not in linear relation, therefore, device operation frequency will demonstrate nonlinear behavior when temperature is derated. The relations of frequency and power dissipation vs. temperature in this case are plotted in Fig.7.3, which shows a minimum frequency value at temperature  $120^\circ C$  and therefore demonstrates a different behavior from that of long channel devices.

Temperature derating behavior around these maximum or minimum frequency values has interesting implication in the process of derating temperature for reliabil-

ity. Simulation identifies relatively flat regions around these extreme value points, so temperature derating within these flat regions will cause little variations in circuit speed and power consumption, which simplifies performance and reliability tradeoffs.

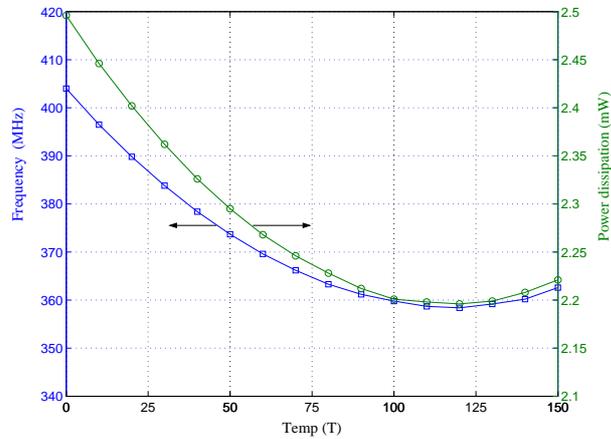


Figure 7.3: The derating curves of frequency and power dissipation vs. temperature. Operation frequency and power dissipation follow nonlinear trends when temperature is derated and simulation shows a minimum frequency value at temperature around  $120^{\circ}C$

The above analysis formulates a guideline for effectively derating temperature for the sake of reliability improvement. The flat region of temperature derating curves must be properly identified, otherwise, short channel devices may not obtain lifetime enhancement even though temperature is derated, or long channel devices may lose the potential to gain lifetime extension by scaling temperature without sacrificing performance.

Threshold voltage, saturation velocity and parasitic drain and source resistances are other important parameters that are sensitive to temperature. Threshold

voltage  $V_t$  increases as temperature decreases due to the shifts of Fermi level and bandgap energy. Saturation velocity  $v_{sat}$  is determined by the critical field and carrier effective mobility  $\mu_{eff}$  thereby also varying with temperature. Although  $\mu_{eff}$  has complicated temperature dependency,  $v_{sat}$  is actually a weak function of temperature and usually demonstrates a simple dependence on temperature:  $v_{sat}$  decreases as temperature increases [146]. Parasitic drain/source series resistance  $R_{ds}$  consists of contact resistance, drain and source diffusion sheet resistance, and spreading resistance resulting from current crowding at the edge of the inversion layer. In the BSIM3v3 model,  $V_t$ ,  $v_{sat}$  and  $R_{ds}$  are all modeled with linear relations to temperature [147].

Derating temperature alone does not influence device performance as much as derating voltage, but reducing temperature and voltage together will produce an order of magnitude reliability improvement. This significant improvement results from the modification of device junction temperature  $T_j$ , which is dependent on the power dissipation  $P_D$ , the ambient temperature  $T_a$  and the thermal impedance  $\theta_{ja}$ :

$$T_j = \theta_{ja}P_D + T_a \quad (7.4)$$

The dependence of  $T_j$  on  $V_{DD}$  is given by:

$$T_j = T_a - \frac{V_{DD}(V_{DD} - V_t)^2(T_a - T_j^0)}{V_{DD}^0(V_{DD}^0 - V_t)^2} \quad (7.5)$$

where  $V_{DD}^0$  and  $T_j^0$  denote normal operating values for voltage and junction temperature,  $V_{DD}$  and  $T_j$  represent derated values for voltage and junction temperature,  $V_t$  is threshold voltage, and  $T_a$  is the ambient temperature. Each of these parameters can be controlled in circuit design [148]. Temperature derating does provide

an alternative to improve device reliability, however, the above temperature derating behaviors are only valid within the temperature range of  $-50^{\circ}C$  to  $150^{\circ}C$  due to the SPICE model limitation. Beyond this range, complicated scattering mechanisms start to dominate and significantly change the temperature behaviors of low  $V_{DD}$  devices, consequently, additional temperature derating model will be required to characterize any new temperature behaviors [147].

### 7.3.3 Voltage Transfer Analysis

Digital integrated circuits consist of various kinds of interconnected logic gates, and the voltage signals are always contaminated by noise. In order to characterize the noise tolerance or immunity of a circuit to undesired external perturbations, designers normally need to explore and properly set the noise margin parameter which is the difference of equivalent voltage levels between output and input of consecutive gates. Noise magnitude must be within noise margin to make logic gates work at correct input and output voltage levels. There are two noise margin parameters:  $NM_L = V_{IL} - V_{OL}$  for low signal levels, and  $NM_H = V_{OH} - V_{IH}$  for high signal levels, where  $V_{IL}$  is input low voltage,  $V_{IH}$  is input high voltage,  $V_{OL}$  is output low voltage and  $V_{OH}$  is output high voltage. These parameters characterize the DC input-output voltage behaviors and determine the circuit noise tolerance to external signal perturbations. Setting proper values for these noise margins is a basic design consideration for realizing intended functions and enabling correct voltage derating.

The simulation results for the two noise margin parameters  $NM_L$  and  $NM_H$  vs.  $V_{DD}$  are plotted in Fig.7.4, which shows that over the voltage derating range of 4.0V to 1.2V,  $NM_L$  and  $NM_H$  approximately decrease linearly with  $V_{DD}$ . Therefore, derating does not change the ratios of noise margin to voltage.

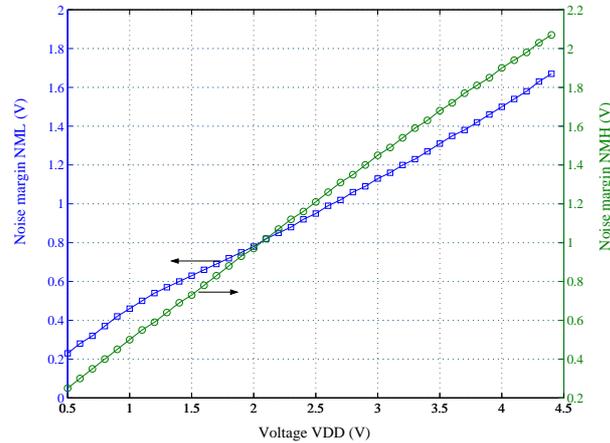


Figure 7.4: The simulation results for  $NM_L$  and  $NM_H$  vs.  $V_{DD}$ . Over the voltage derating range of 4.0V to 1.2V,  $NM_L$  and  $NM_H$  approximately decrease linearly with  $V_{DD}$ .

Sufficient noise margin is very important for a circuit in severe environments where noise can corrupt the circuit signals. Fig.7.4 shows that when  $V_{DD}$  is very small, noise margins will decrease to very low levels. Therefore, in low power applications where noise is ubiquitous, noise margin may impose lower limits on voltage derating. Nevertheless, the frequency can be decreased more than what is required only by voltage reduction to reduce the noise sensitivity, and a derated device can have greater noise tolerance than its full performance specification. Fig.7.5 is the plot of DC VTC under different input voltage dynamic range (from 0.5V to 4.5V). For an ideal CMOS inverter, the output dynamic range is from 0 to  $V_{DD}$ . When

$V_{DD}$  scales down, it is obvious that the width of uncertain region, i.e. transition region, of VTC reduces proportionally. Reducing the width of uncertain region is one of the most important design objectives for lowering power and boosting speed, however, there exists a limit for excessively reducing the width of this transition region due to the MOSFET threshold voltage requirements.

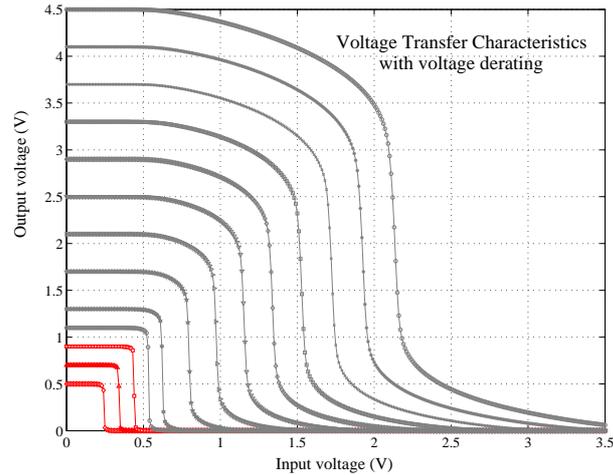


Figure 7.5: DC VTC curves at different power supply voltage (from 0.5V to 4.5V). When  $V_{DD}$  scales down, the transition region of VTC reduces proportionally. When  $V_{DD}$  is lower than 0.8V, the transition region disappears, and the VTC exhibits a hysteresis behavior.

Fig.7.5 shows when  $V_{DD}$  is lower than 0.8V, the transition region disappears, and  $V_{IL}$  quickly approaches to  $V_{IH}$ . In this case, the inverter will operate with a region in which none of the transistors is conducting. This means the inverter cannot function correctly any longer. In theory, the lower limit for  $V_{DD}$  is bounded by the summation of nMOSFET and pMOSFET threshold voltages:

$$V_{DD}(min) = V_{tn} + |V_{tp}| \quad (7.6)$$

For the  $0.18\mu m$  technology being considered,  $V_{tn} = 0.368V$  and  $V_{tp} = 0.435V$ , and their summation is about  $0.8V$ , so the simulation result conforms to the theory. When  $V_{DD}$  is lower than  $V_{DD}(min)$ , the VTC will contain a cut-off region. The output voltage within this region will maintain its previous state due to the charge preservation at the output node. Thus, the inverter VTC exhibits a hysteresis behavior at very low supply voltages [143].

## 7.4 Derating Model and Derating Factor

Careless designers may draw a conclusion from above simulation that the margin is very large to effectively derate voltage and temperature without excessively damaging circuit performance. In fact, designers do not have that much flexibility because they are tied to the published device specifications. When MOS devices go down to deep submicron dimensions, the nominated supply voltages are also lowered to subdue electric fields, e.g. in  $0.13\mu m$  technology, the supply voltage  $V_{DD}$  is as low as  $1.2V$ . According to ITRS 2002 Update, for  $90nm$  technology,  $V_{DD}$  will be even lower than  $1.0V$ . However, threshold voltages have not been scaled down in proportional over the generations, and some aforementioned mechanisms also impose lower limits on  $V_{DD}$ . Consequently, how to properly derate  $V_{DD}$  in valid ranges is not a trivial work. With technology advancement, in-depth understanding of derating behaviors, accurate derating models and practical design guidelines become more and more important for circuit designers.

The idea of derating for reliability finds origin in the principles of Accelerated

Stress Tests (AST) in which devices are over-stressed to precipitate failures within a reasonable short time span and then their reliability parameters are extrapolated back to normal operation conditions. Derating can be treated as a reverse application of AST, and similarly, designers need simple rules to derate devices for desired lifetime improvement without tampering functionality. A new factor, derating factor  $D_f$ , as a counterpart to the acceleration factor in AST is introduced, which is defined as the ratio of the MTTF of a device operating at derated conditions ( $MTTF_d$ ) to its MTTF at rated operation conditions ( $MTTF_0$ ):

$$D_f = \frac{MTTF_d}{MTTF_0} \quad (7.7)$$

$D_f$  can represent the total effect of various wearout mechanisms at circuit level. From the lifetime models presented in the previous chapters, it is easy to obtain the expressions of  $D_f$  for the four wearout mechanisms: HCI, TDDB, NBTI and EM, respectively.

$$D_{f_{HCI}} = \left(\frac{I_{sub}^0}{I_{sub}}\right)^n \exp\left[\frac{E_{aHCI}}{\kappa} \left(\frac{1}{T_j} - \frac{1}{T_j^0}\right)\right] \quad (7.8)$$

$$D_{f_{TDDB}} = \frac{(V_{gs})^{a+bT_j}}{(V_{gs}^0)^{a+bT_j^0}} \exp\left[c\left(\frac{1}{T_j} - \frac{1}{T_j^0}\right) + d\left(\frac{1}{(T_j)^2} - \frac{1}{(T_j^0)^2}\right)\right] \quad (7.9)$$

$$D_{f_{NBTI}} = \left(\frac{V_{gs}^0}{V_{gs}}\right)^{\frac{1}{\beta}} \left[\frac{(1 + 2 \exp(-\frac{E_1}{\kappa T_j^0}))^{-1} + (1 + 2 \exp(-\frac{E_2^0}{\kappa T_j^0}))^{-1}}{(1 + 2 \exp(-\frac{E_1}{\kappa T_j}))^{-1} + (1 + 2 \exp(-\frac{E_2}{\kappa T_j}))^{-1}}\right]^{\frac{1}{\beta}} \quad (7.10)$$

$$D_{f_{EM}} = \left(\frac{J^0}{J}\right)^n \left(\frac{T_j^0}{T_j}\right)^m \exp\left[\frac{E_{aEM}}{\kappa} \left(\frac{1}{T_j} - \frac{1}{T_j^0}\right)\right] \quad (7.11)$$

where  $I_{sub}^0$ ,  $V_{gs}^0$ ,  $J^0$ ,  $E_2^0$  and  $T_j^0$  denote rated operating values for nominated use conditions, while  $I_{sub}$ ,  $V_{gs}$ ,  $J$ ,  $E_2$  and  $T_j$  represent expected derated values. The

above four individual derating factors are related to the total derating factor  $D_f$  with a function  $f_d$ :

$$D_f = f_d(D_{f_{HCI}}, D_{f_{TDDB}}, D_{f_{NBTI}}, D_{f_{EM}}) \quad (7.12)$$

The most important part of a derating model is to determine the function  $f_d$ . Derivation of explicit expression for  $f_d$  is complicated and requires detailed information of circuit architecture and stress conditions. But for a simple analysis, designers can assume that  $D_{f_{HCI}}$ ,  $D_{f_{TDDB}}$ ,  $D_{f_{NBTI}}$  and  $D_{f_{EM}}$  are independent with each other, therefore, within small derating scales,  $f_d$  can be approximated with a linear relation:

$$f_d = C_{HCI}D_{f_{HCI}} + C_{TDDB}D_{f_{TDDB}} + C_{NBTI}D_{f_{NBTI}} + C_{EM}D_{f_{EM}} \quad (7.13)$$

where  $C_{HCI}$ ,  $C_{TDDB}$ ,  $C_{NBTI}$  and  $C_{EM}$  are constants and their values can be determined from experiment or simulation. When the derated condition is the same as the rated condition, there is no derating and the total derating factor  $D_f$  equals to unity:

$$D_f = f_d(1, 1, 1, 1) = 1 \quad (7.14)$$

Equ.(7.14) indicates that the summation of  $C_{HCI}$ ,  $C_{TDDB}$ ,  $C_{NBTI}$  and  $C_{EM}$  always equals to unity for any derating process:

$$C_{HCI} + C_{TDDB} + C_{NBTI} + C_{EM} = 1 \quad (7.15)$$

From equ.(7.8) ~ (7.11), designers can determine  $D_{f_{HCI}}$ ,  $D_{f_{TDDB}}$ ,  $D_{f_{NBTI}}$  and  $D_{f_{EM}}$  under any derated voltage and temperature conditions. If  $C_{HCI}$ ,  $C_{TDDB}$ ,

$C_{NBTI}$  and  $C_{EM}$  are calibrated from simulation or testing work, the overall circuit derating factor  $D_f$  can be predicted from equ.(7.12) and (7.13).

## 7.5 Derating Factor and Simulation

For the purpose of obtaining some knowledge on derating factor and understanding its influence on circuit reliability improvement, the dependence of  $D_f$  on  $V_{DD}$  is simulated and a derating graph is generated within a reasonable scale.

Currently, there is no universally accepted lifetime distribution model for all device wearout mechanisms, but failure information extracted from the customer's maintenance database has been researched and statistical analysis has been performed to obtain information relating to how circuits fail [149]. Overwhelming evidence points to an exponentially distributed failure pattern for aerospace circuits [148]. This prompts an assumption that circuit lifetime distribution is approximately exponential no matter what the lifetime distribution is for each of the device wearout mechanisms.

A simple method to calculate  $C_{HCI}$ ,  $C_{TDDB}$ ,  $C_{NBTI}$  and  $C_{EM}$  starts from an assumption that each wearout mechanism contributes equally to the total derating effect. This is a plausible assumption, otherwise, if any wearout mechanism is more significant than others, designers and manufacturers will develop techniques to attenuate its effect. A good example is the development of LDD structure for suppressing HCI effect. Upon this assumption,  $C_{HCI}$ ,  $C_{TDDB}$ ,  $C_{NBTI}$  and  $C_{EM}$  will

conform to the following ratios:

$$C_{HCI} : C_{TDDB} : C_{NBTI} : C_{EM} = \frac{1}{D_{f_{HCI}}} : \frac{1}{D_{f_{TDDB}}} : \frac{1}{D_{f_{NBTI}}} : \frac{1}{D_{f_{EM}}} \quad (7.16)$$

Combining equ.(7.15) and (7.16), we can easily find the expressions for  $C_{HCI}$ ,  $C_{TDDB}$ ,  $C_{NBTI}$  and  $C_{EM}$  as follows:

$$C_{HCI} = \frac{\frac{1}{D_{f_{HCI}}}}{\frac{1}{D_{f_{HCI}}} + \frac{1}{D_{f_{TDDB}}} + \frac{1}{D_{f_{NBTI}}} + \frac{1}{D_{f_{EM}}}} \quad (7.17)$$

$$C_{TDDB} = \frac{\frac{1}{D_{f_{TDDB}}}}{\frac{1}{D_{f_{HCI}}} + \frac{1}{D_{f_{TDDB}}} + \frac{1}{D_{f_{NBTI}}} + \frac{1}{D_{f_{EM}}}} \quad (7.18)$$

$$C_{NBTI} = \frac{\frac{1}{D_{f_{NBTI}}}}{\frac{1}{D_{f_{HCI}}} + \frac{1}{D_{f_{TDDB}}} + \frac{1}{D_{f_{NBTI}}} + \frac{1}{D_{f_{EM}}}} \quad (7.19)$$

$$C_{EM} = \frac{\frac{1}{D_{f_{EM}}}}{\frac{1}{D_{f_{HCI}}} + \frac{1}{D_{f_{TDDB}}} + \frac{1}{D_{f_{NBTI}}} + \frac{1}{D_{f_{EM}}}} \quad (7.20)$$

Subbing equ.(7.17) ~ (7.20) into (7.12) and (7.13), we end up with a very simple  $D_f$  model:

$$D_f = \frac{4}{\frac{1}{D_{f_{HCI}}} + \frac{1}{D_{f_{TDDB}}} + \frac{1}{D_{f_{NBTI}}} + \frac{1}{D_{f_{EM}}}} \quad (7.21)$$

The voltage derating trends governed by this simple  $D_f$  model is simulated with typical model parameters from the  $0.18\mu m$  technology.  $V_{DD}$  is derated within the range [100% ~ 80%] of its rated value, i.e.  $V_{DD}^0 = 3.3V$ . Fig.7.6 is the plotting of the relation between  $D_f$  and  $V_{DD}/V_{DD}^0$ , which shows that within the derating range, the dependency of  $D_f$  on  $V_{DD}$ , after normalized to  $V_{DD}^0$ , is in exponential relation. Fig.7.6 also indicates that the variations of individual derating factors are

quite different, up to 3 orders of magnitude in difference. Another derating factor for the lower rated voltage  $V_{DD}^0 = 1.8V$  is also plotted in Fig.7.6. These two derating factors at different rated voltages almost follow the same trend, which reveals that no matter what the rated voltage is, if voltage is derated to the same ratio, the reliability gain is nearly the same. This is a very important derating guideline. The above derating analyses have been verified by the experimental work in [150].

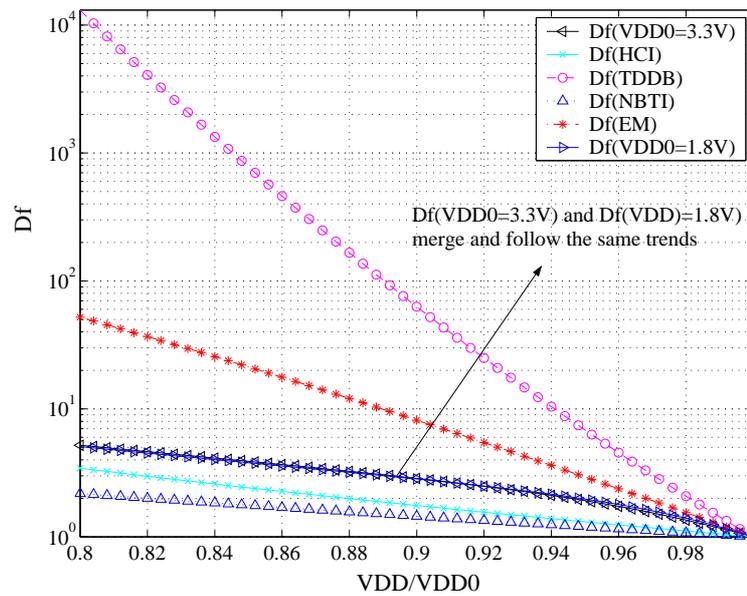


Figure 7.6: Trends of  $D_f$  vs.  $V_{DD}/V_{DD}^0$  with typical model parameters from the  $0.18\mu m$  technology.  $V_{DD}$  is derated within the range [100% ~ 80%] of its rated value  $V_{DD}^0 = 3.3V$ . The trend of  $D_f$  when  $V_{DD}^0 = 1.8V$  is also plotted for comparison.

## 7.6 Conclusion

With the lifetime and circuit models of various wearout mechanisms developed in the previous chapters, MaCRO is capable of performing many distinct reliability

analysis functions including failure rate calculation, reliability trend prediction, derating modeling and circuit SPICE reliability simulation. The MaCRO flows for failure rate calculation and reliability trend prediction are straightforward and they are illustrated in Appendix B. The application of MaCRO models in circuit SPICE reliability simulation is presented in the next chapter. This chapter focuses on MaCRO derating modeling for reliability improvement. From the simulation work of a 17-stage  $0.18\mu m$  CMOS ring oscillator, the voltage and temperature derating behaviors are systematically investigated, and a simple derating factor model is developed. A series of derating design guidelines are formulated during the development of this derating factor model. Circuit designers, as well as system developers, can use these guidelines, or even explore new rules with this simple model, to properly derate devices for reliability improvement for long life applications.

## Chapter 8

### SRAM Reliability Simulation and Analysis

#### 8.1 Introduction

The lifetime models and circuit models for HCI/TDDB/NBTI failure mechanisms as well as the overall reliability simulation algorithms in MaCRO have been presented in the previous chapters. This chapter is an illustrative case study for the purpose of demonstrating how to apply MaCRO models and algorithms to circuit reliability simulation, analysis and improvement.

The most common circuit structures used in exemplary reliability simulations are ring oscillator, differential amplifier and SRAM. Compared with the other two circuits, SRAM includes many typical subcircuits such as cross-connected 6-T memory cell, precharge, peripheral control logic and sense amplifier. The magnitude of MOSFET's wearout mechanisms and their effects on circuit performance and functionality depend on the types of circuits involved [151]. Moreover, for a typical SoC circuit, SRAM occupies more than 40% of the chip area [152]. The ever-increasing integration of SRAM in embedded SoC indicates that the reliability of modern VLSI systems depends on the reliability of on-chip memories [154]. Therefore, SRAM is selected in this case study as a vehicle to show the applicability of MaCRO models

and algorithms in circuit reliability simulation and analysis.

## 8.2 SRAM Circuit Design and Simulation

In order to simplify the circuit structure, reduce reliability simulation complexity and magnify the effects of each failure mechanism on circuit operation, only one bit SRAM cell and its operation control functions are implemented. The address decoder and complex timing control subcircuits are intentionally omitted. The SRAM circuit chosen for this consideration includes one 6-T cell, precharge, read/write control and sense amplifier. The SRAM structural block diagram is shown in Fig.8.1. The detailed structure and function of each block are introduced in this section. The overall circuit is implemented with a commercial  $0.25\mu m$  technology with gate oxide thickness  $5.7nm$  and power supply voltage  $2.5V$ .

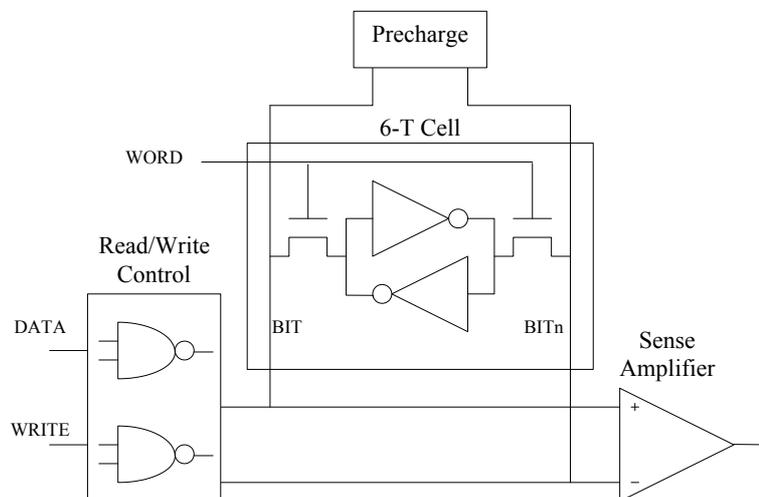


Figure 8.1: The one bit SRAM structural block diagram. The circuit consists of one bit 6-T cell, read/write control logic and output sense amplifier.

The most important functional block in Fig.8.1 is the one bit 6-T SRAM cell which consists of a pair of cross-connected inverters and two nMOSFET pass transistors. The schematic of the SRAM cell is shown in Fig.8.2. Transistors  $M1 \sim M4$  form a regenerative structure for storing a single bit “1” or “0” at the node “Store” depending on the differential voltages of BIT/BITn lines during write cycles. The WORD line controls the two pass transistors  $M5$  and  $M6$  and enables charging/discharging paths between the nodes Store/Storen and BIT/BITn lines during read/write cycles. The cell transfer ratio of pass transistor to pull-down nMOSFET widths (i.e. width ratio of  $M5$  to  $M1$ , and  $M6$  to  $M2$ ) is designed to 1. The proper value of this ratio is important for cell stability during read operation [152]. The two transmission gates (consisting of  $M41 \sim M44$ ) provide bidirectional paths and connect BIT/BITn lines to write control circuit during write operation, and to sense amplifier during read operation.

The function of precharge circuit is pre-charging BIT and BITn lines to the same level before each read and write operation. The schematic of precharge circuit is shown in Fig.8.3. When PRE signal is high,  $M21 \sim M25$  turn on, equalizing and charging up BIT/BITn lines to the same voltage level  $V_{DD} - 2V_t$ . Because nMOSFET threshold voltage  $V_t = 0.65V$ , the pre-charge voltage level is approximately set to the middle of  $V_{DD}$ , which avoids full rail-to-rail signal transitions in subsequent read/write operation, thereby improving circuit operation speed. The high speed transition of PRE on  $M21 \sim M25$  may introduce charge injection effects on BIT/BITn lines. These transient charges will increase voltage overshooting and reduce cell stability. For high-speed high-volume SRAM circuits in which node ca-



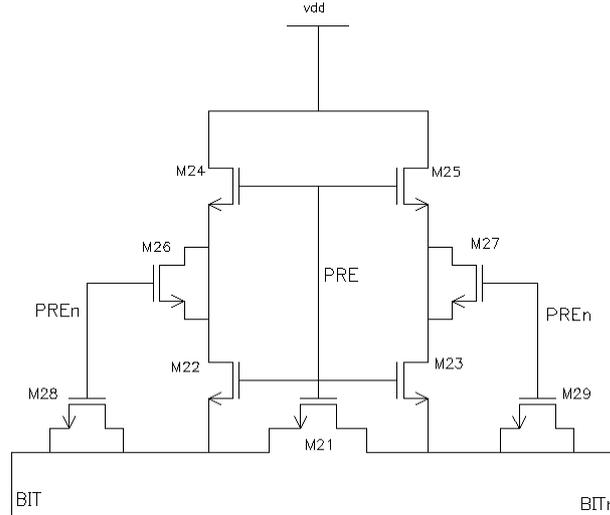


Figure 8.3: Schematic of the precharge circuit. BIT/BITn lines are pre-charged to the same voltage level before each read and write operation.  $M26 \sim M29$  are included for reducing transient charge injection effects.

Latch type sense amplifier rather than current mirror amplifier is selected due to the small node capacitances and large voltage swings of BIT/BITn lines. READ signal applies to  $M55$  and  $M60$  and controls the read operation. If READ signal is high, the latch amplifier, consisting of  $M51 \sim M55$ , quickly pulls BIT/BITn apart in reverse directions to the full digital levels.  $M56 \sim M59$  form the output buffer and help to generate smooth rail-to-rail output signals. The overall schematic of the sense amplifier circuit is illustrated in Fig.8.5.

The function of the SRAM is simulated in SPICE to perform a set of sequential “write 0, read 0, write 1, read 1” operations. The duration of each operation cycle is  $2ns$ , and the circuit is simulated for  $8ns$  with operation speed of  $500MHz$ . The timing of input signals is given in Fig.8.6.

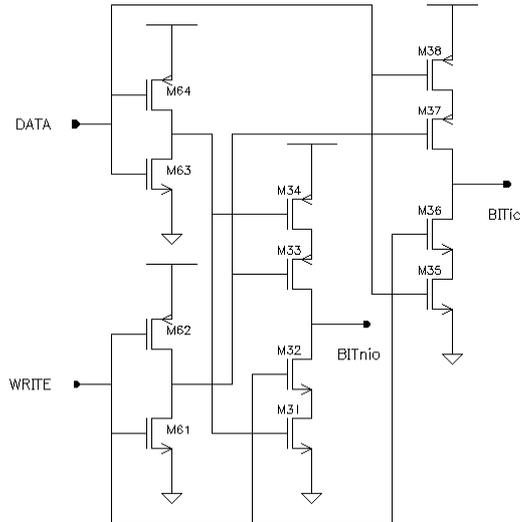


Figure 8.4: Schematic of the write control circuit. WRITE signal controls the connections of DATA line and BITio/BITnio lines. BITio/BITnio lines are connected to BIT/BITn by the two transmission gates ( $M41 \sim M44$ ).

The SPICE simulation results are shown in Fig.8.7 in which (a) demonstrates precharging states and swings of BIT/BITn signals during read/write operations, (b) indicates SRAM cell state stored at Store/Storen nodes, (c) shows results of the two write operations, and (d) shows results of the two read operations. These simulation waveforms illustrate the SRAM operation process: within  $1 \sim 2ns$ , “0” on the DATA line is written into the SRAM cell, within  $3 \sim 4ns$ , “0” state stored in the SRAM cell is read out to the output data line DATAO, within  $5 \sim 6ns$ , “1” on the DATA line is written into the SRAM cell, and in  $7 \sim 8ns$ , “1” state stored in the SRAM cell is read out to DATAO. These timing relations will be compared later with MaCRO reliability simulation results after SRAM experiencing HCI/TDDDB/NBTI stresses.

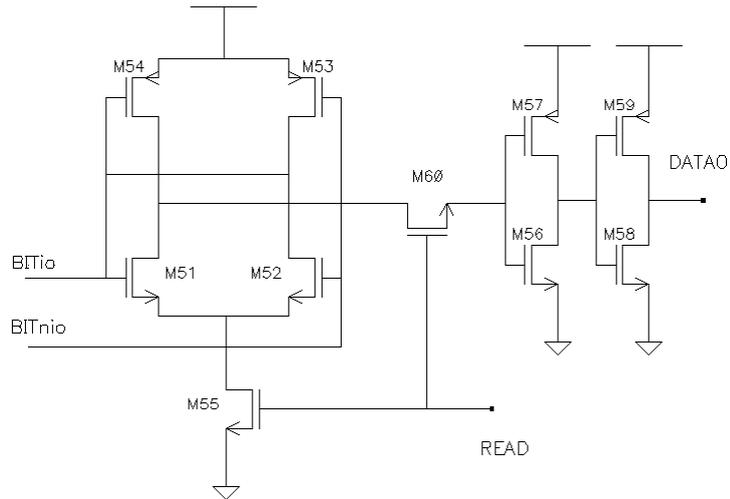


Figure 8.5: Schematic of the sense amplifier. READ signal controls the operation of the latch amplifier and the connection between BIT/BITn and the output. The latch amplifier magnifies BIT/BITn line swings to full digital levels.

### 8.3 Preview of SRAM Failure Behaviors

For the sake of facilitating the understanding of MaCRO reliability simulation results, a brief overview of SRAM reliability behaviors and failure effects presented in literature is given in this section.

The main effects of HCI on device electrical characteristics are threshold voltage drift and transconductance ( $g_m$ ) degradation. Pass transistors in an SRAM cell receive more severe damages because of bidirectional HCI stresses. This is proved by the following MaCRO simulation. The  $g_m$  degradation of these pass transistors gradually reduces the driving capability and cell transfer ratio [151] and increases access time after long term operation [155, 156]. The physical origin of this enhanced HCI damage on pass transistors is explained in [157]. Sense amplifier also suffers from significant HCI stress, which results in increased input offset voltage [158] and

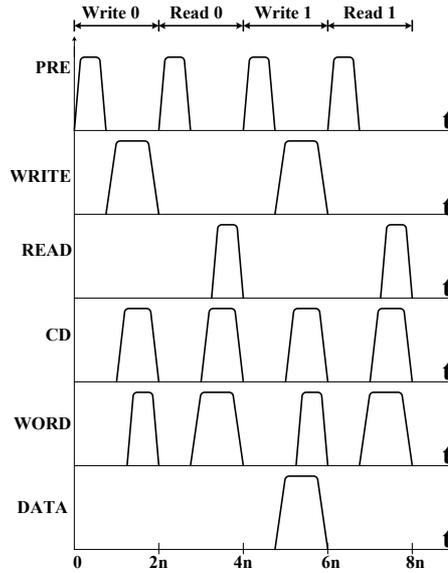


Figure 8.6: SRAM SPICE simulation stimuli. PRE exerts before each read/write operation. CD signal enables the transmission gates  $M41 \sim M42$  and WORD signal enables the pass transistors  $M5 \sim M6$  during each read/write operation. The “0” or “1” is available on DATA line during each write operation.

decreased drain output resistance and small-signal voltage gain [159].

TDDDB has the most deleterious effects on SRAM cell stability. There are only four topologically distinct oxide breakdown locations in the SRAM cell shown in Fig.8.2: Store-to-Storen, Store-to- $V_{DD}$ , Store-to- $gnd$ , and gate-to-diffusion of pass transistors. Any other possible oxide breakdown location is completely equivalent to one of these categories [152]. Store-to-Storen breakdown and gate-to-diffusion breakdown of pass transistors reduce BIT/BITn differential voltage and output swing, whereas breakdowns at Store-to- $V_{DD}$  and Store-to- $gnd$  increase leakage current at the opposite transistors and degrade cell stability and Static Noise Margin (SNM) [160]. The leakage currents of  $20 \sim 50\mu A$  at the nMOSFET source can result in a

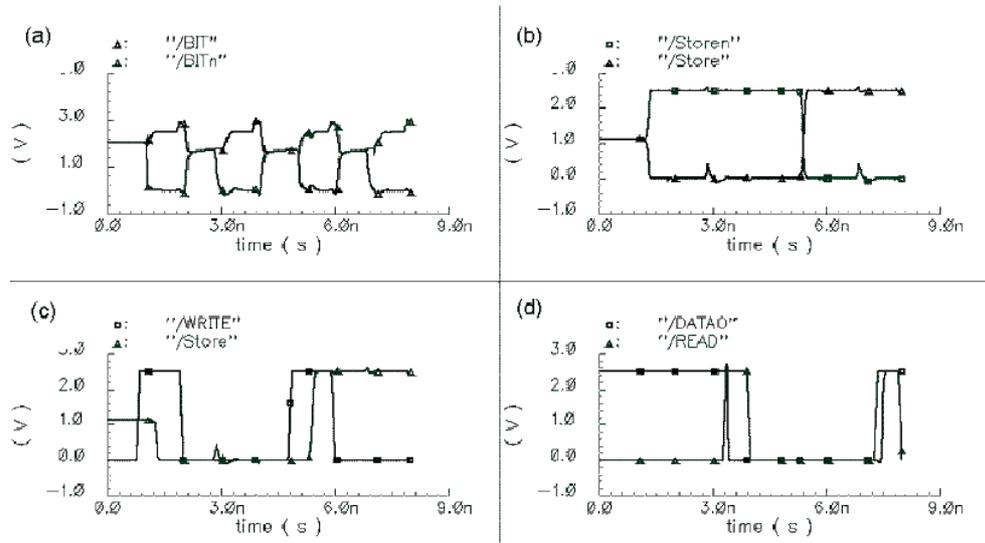


Figure 8.7: SRAM SPICE simulation results. (a) shows waveforms of BIT/BITn signals, (b) shows SRAM cell state signals Store/Storen, (c) is write operation result, and (d) is read operation result.

50% reduction in SNM [161, 162]. Most SRAM cells become unstable without sufficient SNM [163]. A thorough investigation of different gate oxide breakdown effects on SRAM subcircuits is presented in [164, 165].

NBTI leads to device mismatches in the SRAM cell and input offset voltages in the sense amplifier. The SNM degradation due to NBTI increases as  $V_{DD}$  decreases [166]. Experimental work of an operational amplifier to end-of-life degradation indicates little change in output characteristics, suggesting that pMOSFET NBTI-induced device mismatch is not the fundamental reason for circuit failures [167]. This conclusion is also supported by the following MaCRO reliability simulation results.

## 8.4 Device Lifetime Calculation

The lifetime model for each failure mechanism (HCI/TDDB/NBTI) is introduced in the previous chapters. These lifetime equations are recapitulated here for convenience:

$$t_f(HCI) = A_{HCI} \left( \frac{I_{sub}}{W} \right)^{-n} \exp\left( \frac{E_{aHCI}}{\kappa T} \right) \quad (8.1)$$

$$t_f(TDDB) = A_{TDDB} \left( \frac{1}{A} \right)^{\frac{1}{\beta}} F^{\frac{1}{\beta}} V_{gs}^{a+bT} \exp\left( \frac{c}{T} + \frac{d}{T^2} \right) \quad (8.2)$$

$$t_f(NBTI) = A_{NBTI} V_{gs}^{-\frac{1}{\beta}} \left[ \frac{1}{1 + 2 \exp\left(-\frac{E_1}{\kappa T}\right)} + \frac{1}{1 + 2 \exp\left(-\frac{E_2}{\kappa T}\right)} \right]^{-\frac{1}{\beta}} \quad (8.3)$$

If all the model parameters are determined from device testing work, based on SPICE simulation results, designers can calculate device lifetime for each failure mechanism at any use conditions. However, from the perspective of circuit functionality, absolute value of device lifetime is not of primary interest. The main purpose of lifetime calculation is for identification of most weakest and damaged devices, so it is only required to calculate relative lifetime (i.e. normalized lifetime) for each device by lumping all common model parameters into a single factor. Based on this concept, equ.(8.1) ~ (8.3) can be rewritten to the following simplified forms:

$$t_f(HCI) = \tau_1 \left( \frac{I_{sub}}{W} \right)^{-n} \quad (8.4)$$

$$t_f(TDDB) = \tau_2 \left( \frac{1}{W} \right)^{\frac{1}{\beta}} V_{gs}^{a+bT} \quad (8.5)$$

$$t_f(NBTI) = \tau_3 V_{gs}^{-\frac{1}{\beta}} \left[ E'_1 + \frac{1}{1 + 2 \exp\left(-E_2/\kappa T\right)} \right]^{-\frac{1}{\beta}} \quad (8.6)$$

where  $\tau_1 \sim \tau_3$  are the lumped factors and defined as benchmarks for normalized lifetimes,  $W$  is the channel width,  $E'_1$  is a process dependent constant. In deriving equ.(8.4)  $\sim$  (8.6), device junction temperature and the ambient temperature are not differentiated. The temperature effects of various failure mechanisms are discussed in [153] and the method to model device junction temperature with respect to device power dissipation and ambient temperature is given in [154].

In normalized lifetime calculation process, it is unnecessary to characterize  $\tau_1 \sim \tau_3$  factors because they are common to all devices in the same circuit. This reduces the number of model parameters and obviously simplifies parameter testing and extraction work. In equ.(8.4)  $\sim$  (8.6),  $I_{sub}$ ,  $V_{gs}$  and  $E_2$  can be predicted from SPICE simulation. After obtaining the reduced set of model parameters necessary to equ.(8.4)  $\sim$  (8.6), designers can easily calculate device normalized lifetimes for each failure mechanism. The lifetime results are shown in Fig.8.8, in which horizontal axis denotes transistor's index (e.g. "1" represents "M1"), and vertical axis denotes lifetime value normalized to  $\tau_1 \sim \tau_3$ , respectively (e.g. for HCI:  $t_f(M1) = 4.2893\tau_1$ ). Compared with other devices,  $M33, M34, M37, M41, M43$  have very large NBTI lifetimes. In order to show details of other devices' relatively smaller lifetime values, normalized lifetime values of these transistors are not drawn in scale in (c) of Fig.8.8.

The following trends can be easily observed from inspecting Fig.8.8. For HCI effect, pass transistors generally experience more damage due to bidirectional stresses and more frequent switching operations, shown by  $M5, M6, M21, M42, M44$ ; nMOSFETs in inverters suffer from less HCI stress, shown by  $M35, M56, M63$ ; in stacked inverters, nMOSFETs on the top receives more HCI damage, shown by

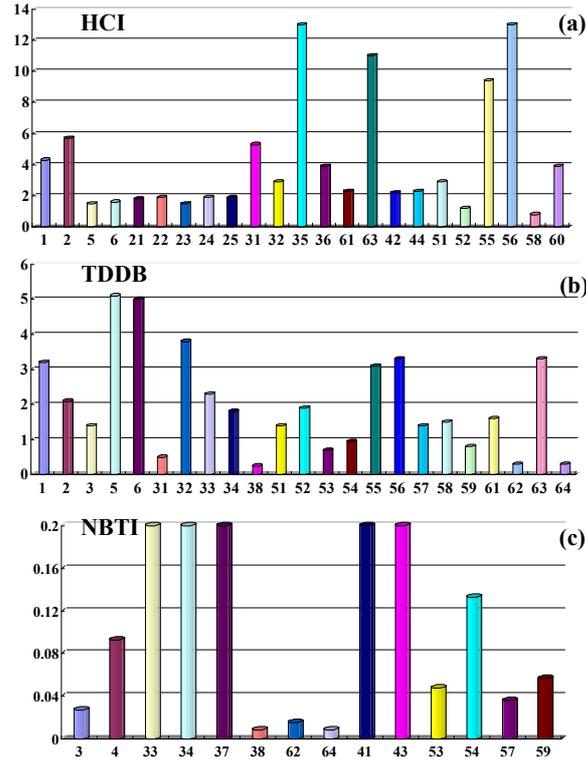


Figure 8.8: Device lifetime calculation results for the three failure mechanisms: (a) HCI, (b) TDDB, and (c) NBTI. The horizontal axis denotes device's index, and vertical axis denotes lifetime value normalized to  $\tau_1 \sim \tau_3$ , respectively.

comparisons of  $M31$  to  $M32$  and  $M35$  to  $M36$ , respectively; sense amplifier is sensitive to HCI because distinct HCI damages on  $M51$  and  $M52$  lead to increased device mismatches and input offset voltages. For TDDB effect, pMOSFET is easier to suffer from TDDB due to its relatively larger channel area, and area scaling has significant effect on device lifetimes, shown by  $M62$ ,  $M64$  whose channel widths are very large:  $12\mu m$ . For NBTI effect, pMOSFETs in latch structure receive more imbalanced NBTI damages, which also leads to increased device mismatches and input offset voltages, shown by  $M3$ ,  $M4$  and  $M53$ ,  $M54$ .

It is easy to identify the most damaged transistors for each failure mechanism from Fig.8.8. For HCI,  $M5$ ,  $M6$ ,  $M52$ ,  $M58$  are the most damaged transistors.  $M58$  has the shortest lifetime, however, it can be excluded after a careful analysis. In the initial schematic, the two stages of inverters after sense amplifier were designed with the sizing ratio of 1. If scaling up the channel widths of  $M58$  and  $M59$  and increasing the sizing ratio to 3, the lifetime of  $M58$  increases from  $0.84\tau_1$  to  $7.79\tau_1$ . The reason for this significant improvement is the reduction in inverter transition delay after proper sizing of inverter chain as shown in Fig.8.9. Proper inverter sizing improves both transition speed and device lifetime with the penalties of larger chip area and loading to neighboring gates, therefore, circuit designers need to perform detailed lifetime calculation and functional simulation to make appropriate tradeoffs.

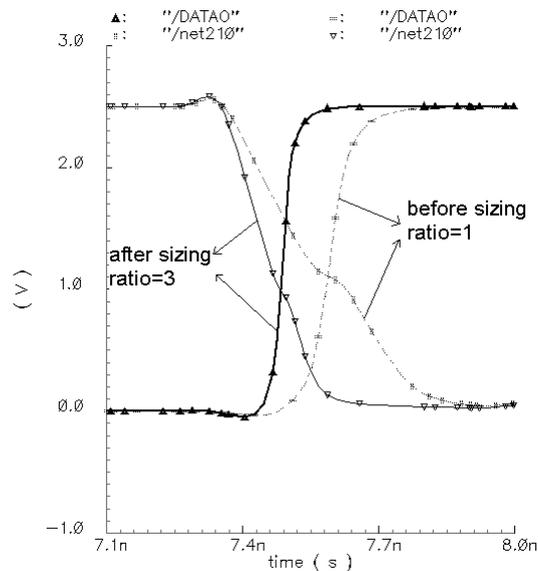


Figure 8.9: Comparison of transition delay of  $M58$  before and after inverter sizing. Proper sizing significantly reduces dynamic switching delay, thereby suppressing HCI effect.  $W_n = 0.6\mu m$  before sizing, and  $W_n = 1.8\mu m$  after sizing.

For TDDB,  $M3, M31, M38, M53, M62, M64$  are the most damaged transistors.  $M62, M64$  are pMOSFETs in write control logic subcircuit, their channel widths are designed very large to quickly generate inverse signals of WRITE and DATA. Their widths can be properly scaled down to improve lifetime, therefore, it is unnecessary to include them in the weakest device list.  $M3$  is included because it is within the SRAM cell and its oxide breakdown has significant effect on SRAM operation. All transistors in precharge circuit ( $M21 \sim M29$ ) are not selected because during all operation periods, their gate-to-source/drain voltages are very small.

For NBTI,  $M3, M38, M53$  are selected as the most damaged transistors. Although lifetimes of  $M62, M64, M57, M59$  are comparable to those of being selected, based on the same reason given above, they are not included in the weakest device list.

In summary, the selected most damaged devices for each failure mechanism are: HCI— $M5, M6, M52$ , TDDB— $M3, M31, M38, M53$ , and NBTI— $M3, M38, M53$ . These transistors will be substituted with corresponding circuit models in the following SPICE simulation.

## 8.5 SPICE Reliability Simulation with Circuit Models

The model equations and methods to determine circuit model parameters have been presented in the previous chapters. Most of these model parameters are time dependent, therefore, SPICE simulation with these circuit models has to be performed several times to pinpoint the time at which the circuit function fails. The

most effective way to find this failure time is by a three-step progressive process: first, only consider HCI failure electrical models and find the circuit HCI lifetime  $T_a$ ; then, add on TDDB electrical models, simulation circuit operation at times shorter than  $T_a$  and find the corresponding circuit HCI+TDDB failure lifetime  $T_b$  ( $T_b \leq T_a$ ); finally, include all failure electrical models and find the circuit failure lifetime  $T_c$  ( $T_c \leq T_b$ ) at which the circuit cannot maintain correct operations. In this step-by-step process, circuit failure behaviors and response due to each failure mechanism can be efficiently characterized. The detailed algorithm of this process is given in Chapter 2. The SRAM reliability simulation and analysis are performed according to this three-step process.

### 8.5.1 HCI

There is only one parameter in HCI circuit model:  $\Delta R_d$ , which characterizes drain current reduction due to mobility degradation resulting from HCI-induced interface charge and oxide charge.  $\Delta R_d$  values of  $M5, M6, M52$  at different stress times are plotted in Fig.8.10. These HCI-induced series parasitic resistances are not in simple logarithmic relation to stress time  $t$  because horizontal axis is not drawn in linear scale.  $M5, M6$  receive bidirectional HCI stresses, consequently, each of them has two resistances  $\Delta R_{d1}$  and  $\Delta R_{d2}$  associated with drain and source, respectively.

The SRAM circuit with these HCI-induced  $\Delta R_d$  elements is simulated at different stress times to check its functionality. Fig.8.11 shows the waveforms of the SRAM cell state (i.e. Store signal) and output state (i.e. DATAO signal) after

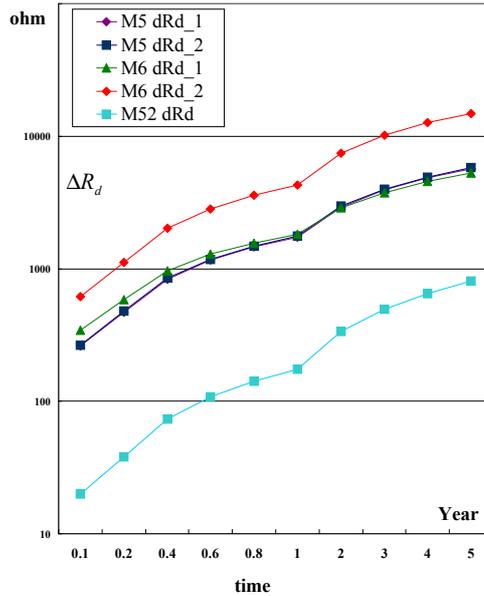


Figure 8.10:  $\Delta R_d$  values of  $M5$ ,  $M6$ ,  $M52$  at different stress times. The unit of horizontal axis is time in years, the vertical axis is in logarithmic scale and in unit Ohm.

different stress times. It indicates that the SRAM circuit operates correctly until 0.8 year, and fails at 1 year, at which the Store signal does not switch as expected during the “write 1” cycle. The gradual degradation of Store signal is clearly shown in Fig.8.11. The quicker corruption of Store signal than that of DATA0 implies that malfunction of this SRAM circuit mainly results from HCI damage of  $M5$ ,  $M6$ , rather than  $M52$ , which verifies other researchers’ work on the relation between pass transistor’s HCI degradation and SRAM cell stability.

A closer look at BIT/BITn and Store/Storen waveforms before and after SRAM cell failure reveals more reliability information. Fig.8.12 compares and shows how these signals corrupt. It is clearly shown that the addition and increasing of

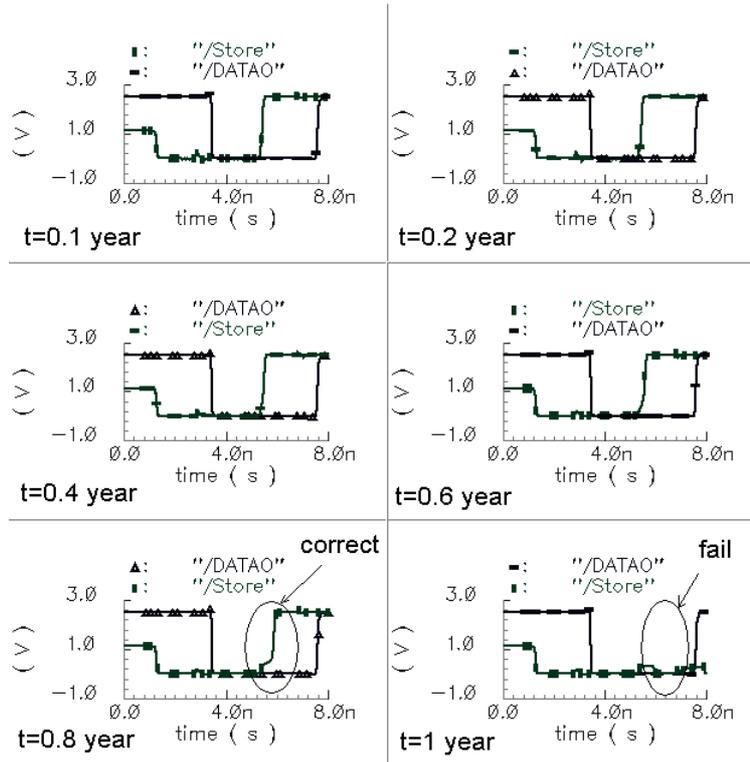


Figure 8.11: The simulated waveforms of the SRAM cell Store signal and output DATA0 signal after different stress times. At  $t = 1$  year, Store signal does not jump to high as expected during the “write 1” cycle indicating failure of SRAM cell.

HCI-induced series resistances in  $M5$  and  $M6$  degrade BIT/BITn signals and reduce cell transfer ratio. As a result, the high BIT line signal at “write 1” cycle cannot be effectively written into SRAM cell. Store/Storen signals cannot switch when a reverse value is being written to the SRAM cell.

From the above SPICE simulation with HCI circuit models, the HCI lifetime of the SRAM circuit is predicted to be 0.9 year. If considering the effect of duty cycle and assuming that the average access frequency of the SRAM is one full operation per  $1\mu s$  at normal use condition, the predicted 0.9 year corresponds to a circuit HCI lifetime of 112.5 years.

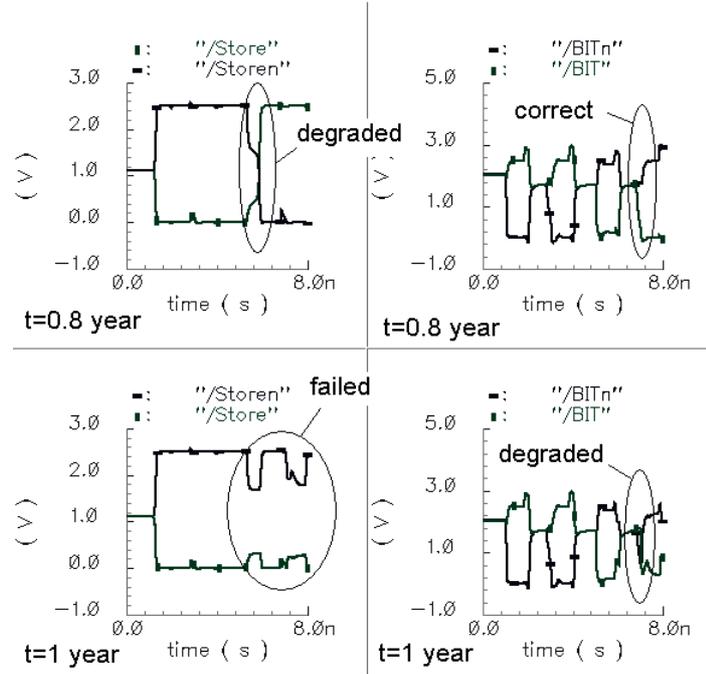


Figure 8.12: The waveforms of SRAM Store/Storen signals and BIT/BITn signals before and after circuit failure. Store/Storen signals do not flip due to the degradation in BIT/BITn signals when a reverse value is being written to the SRAM cell.

### 8.5.2 HCI+TDDB

The second step in the SRAM circuit reliability simulation is the inclusion of both TDDB and HCI circuit models. Only gate-to-channel breakdown is considered and breakdown location is intentionally set to the middle point of the channel. As a result, only one parameter  $I_{ox}$  needs to be characterized for each identified TDDB damaged transistor. The values of  $I_{ox}$  have been calculated as:  $I_{ox}(M3) = -50.719\mu A$ ,  $I_{ox}(M31) = 25.642\mu A$ ,  $I_{ox}(M38) = -18.07\mu A$ , and  $I_{ox}(M53) = -101.05\mu A$ .

The SPICE simulation results when taking into account both HCI and TDDB

effects are illustrated in Fig.8.13. The SRAM circuit survives until 0.4 year but fails to function at 0.6 year.

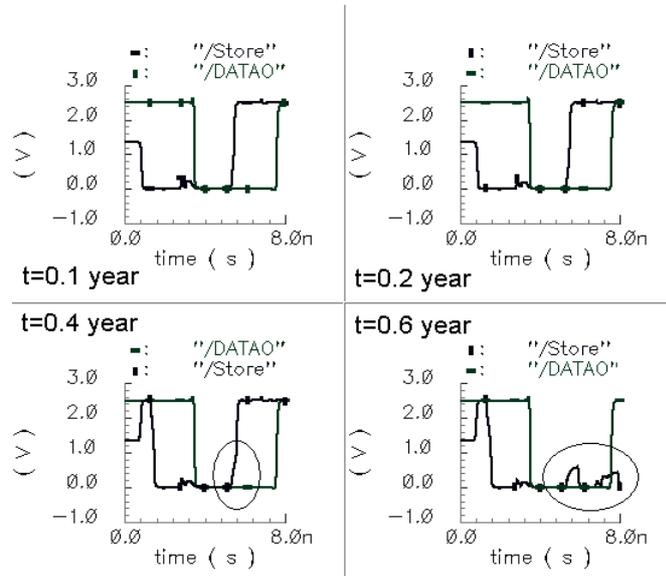


Figure 8.13: The simulated waveforms of the SRAM cell Store signal and output DATA0 signal at different HCI+TDDB stress times. At  $t = 0.6$  year, Store signal does not jump to high during the “write 1” cycle indicating failure of SRAM cell.

The addition of TDDB failure electrical models significantly reduces circuit lifetime. Fig.8.14 shows the interaction between HCI effect and TDDB effect, in which the BIT/BITn and Store/Storen waveforms before and after circuit failure (at 0.4 year and 0.6 year, respectively) are plotted. At 0.6 year, the corruption of Store/Storen signals and the degradation of BIT/BITn signals during the final “write 1, read 1” cycles are very similar to those at 1 year in Fig.8.12, in which only HCI effect is considered. Moreover, if TDDB effect on  $M3$  is disabled at 0.6 year, the circuit function restores and the waveforms without TDDB at 0.6 year are quite similar to the waveforms with TDDB at 0.4 year. These similarities imply

that gate-to-channel breakdown of TDDB accelerates SRAM cell instability but it does not introduce new failure behavior at circuit level. This result is not observed by other researchers because most work on SRAM cell instability analysis does not combine HCI and TDDB effects together and worst case gate-to-diffusion breakdown mode rather than more frequent and less severe gate-to-channel breakdown mode of TDDB is included in those simulation work.

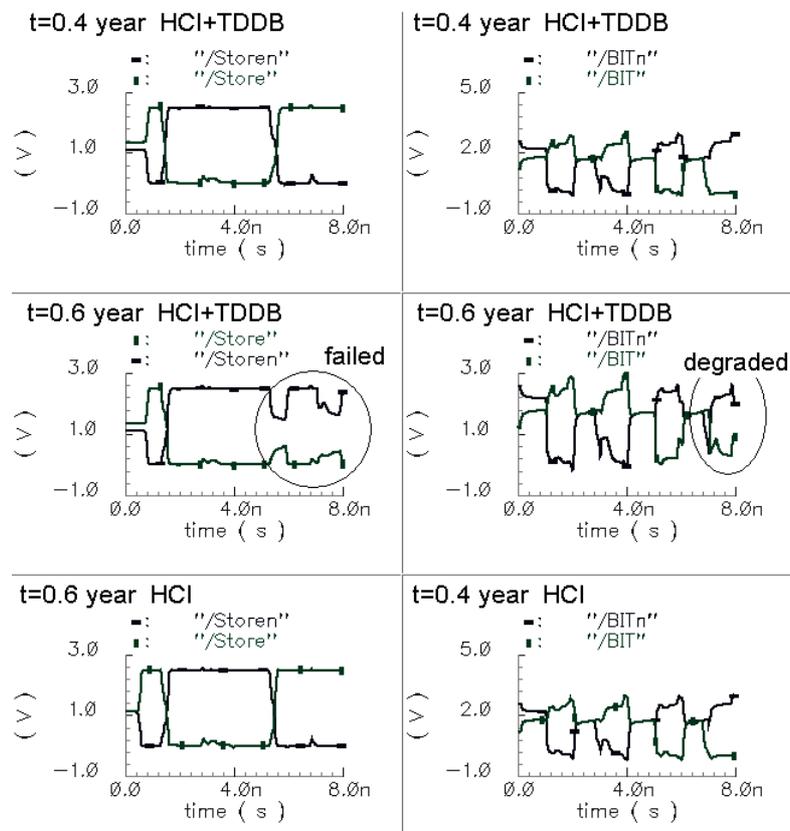


Figure 8.14: The waveforms of the SRAM Store/Store signals and BIT/BITn signals before and after circuit failure.

Besides TDDB of  $M3$  on circuit operation, it is also necessary to investigate TDDB effects of  $M31, M38, M53$  on circuit performance. Simulation proves breakdowns of  $M31$  and  $M38$  (both belong to inverters in write control subcircuit)

have minor effects on SRAM operation, but breakdown of  $M53$  has significant effect. Fig.8.15 shows the TDDB effect of  $M53$  on sense amplifier input signals. The breakdown in  $M53$  provides additional current path between sense amplifier input and  $V_{DD}$  and tends to pull up this input signal. The erratic jumps in the amplifier input signal shown in Fig.8.15 reduce amplifier output stability.

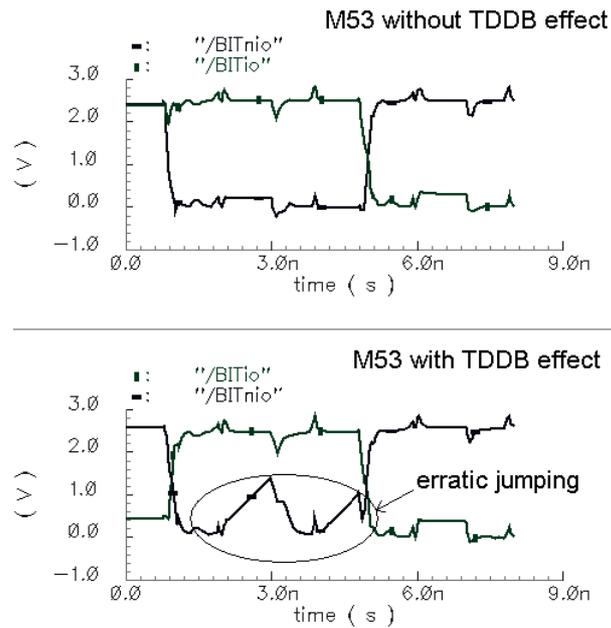


Figure 8.15: The TDDB effect of  $M53$  on sense amplifier output stability. The breakdown in  $M53$  provides additional current path between  $\text{BIT}_{\text{nio}}$  and  $V_{DD}$  and tends to pull up  $\text{BIT}_{\text{nio}}$  when it is at low level in “read 0” and “write 1” cycles.

### 8.5.3 HCI+TDDB+NBTI

The last step is the inclusion of NBTI circuit models.  $M3, M38, M53$ , being identified for suffering most NBTI damage, also receive most TDDB damage, therefore, designers need to properly combine NBTI and TDDB electrical models together

for these pMOSFETs. If they simply add all NBTI failure circuit model elements into TDDB model, the oxide breakdown effect will be overestimated, which results in suppressing or overshooting of SRAM cell state signals (i.e. Sote/Storen), and unexpected jumps of sense amplifier input signals. These negative phenomena are observed in simulation results. The correct TDDB+NBTI failure electrical model for a pMOSFET is illustrated in Fig.8.16.

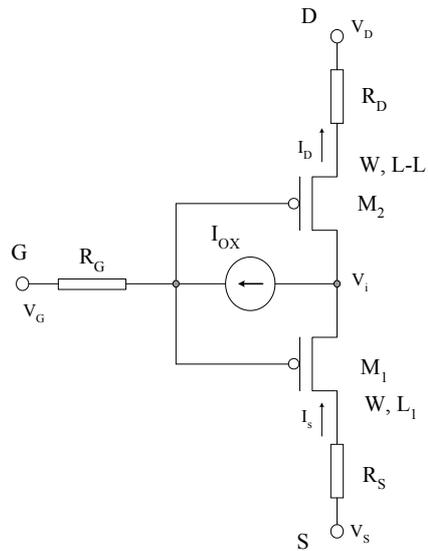


Figure 8.16: The TDDB+NBTI circuit model for a pMOSFET.  $R_G$  and  $I_{ox}$  account for threshold voltage degradation due to NBTI.  $I_{ox}$  and the two split pMOSFETs represent TDDB damage.  $R_D$  and  $R_S$  characterize the resistances in drain and source extensions. They are excluded in this SRAM case study in order to simplify simulation work.

With the previous HCI+TDDB simulation results, it is only required to calculate  $R_G$  for each of  $M3$ ,  $M38$ ,  $M53$  at time 0.4 year. Their values are:  $R_G(M3) = 6.6K\Omega$ ,  $R_G(M38) = 965.4\Omega$ , and  $R_G(M53) = 3.3K\Omega$ . Simulation indicates that

NBTI has relatively weak effects on SRAM cell stability and functionality. Its most obvious influence observed from simulation is that NBTI degrades SRAM cell transition speed. This effect is shown in Fig.8.17 where switching of Store/Storen signals slows down when NBTI model is set in.

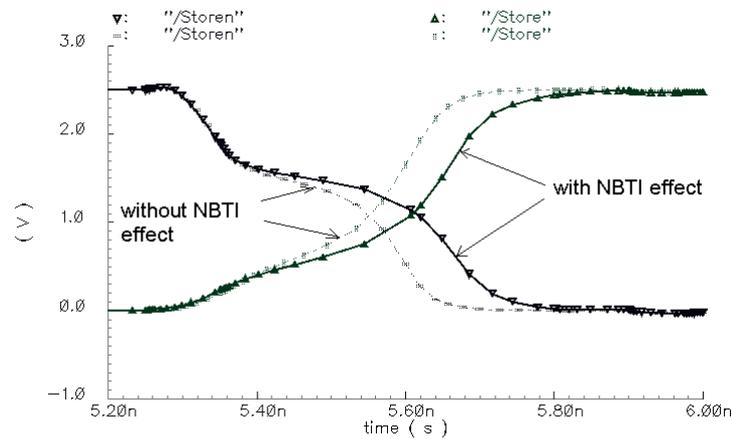


Figure 8.17: The NBTI effects on SRAM cell transition speed. The switching speed of SRAM cell Store/Storen signals degrades when NBTI damage on  $M3$  is considered.

Simulation also shows NBTI has minor effects on functionality of the latch type sense amplifier. The degradation in input signals is very small.

SPICE DC voltage transfer function simulation along the path from BITn line to Storen line encompasses all of the three failure mechanisms (HCI of  $M6$ , TDDB and NBTI of  $M3$ ), therefore, degradation in VTC for BITn-to-Storen at different combinations of these failure mechanisms can reflect their individual influence on SRAM cell stability. These VTC curves, plotted in Fig.8.18, indicate that HCI and TDDB have reverse effects on VTC drift, while NBTI has no observable effects.

SNM is the most important factor in SRAM circuit reliability analysis. Based

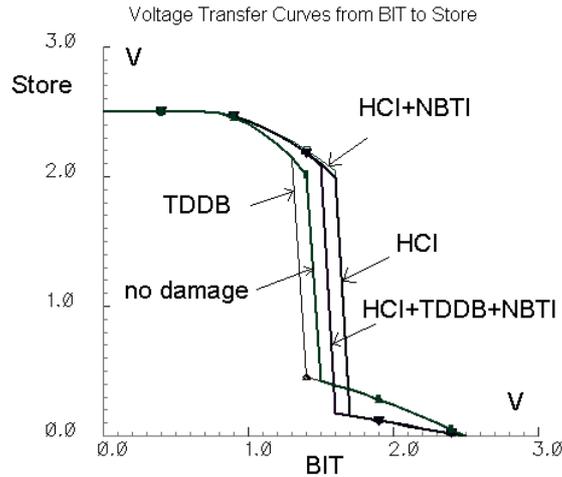


Figure 8.18: Voltage transfer curves of BITn-to-Storen for different combinations of failure mechanisms. From left to right, the curves represent effects of TDDB, no damage, HCI+TDDB+NBTI, HCI, and HCI+NBTI, respectively. NBTI has negligible effect on SRAM cell stability.

on SPICE DC transfer analysis, the SNM butterfly plots for various combinations of the failure mechanisms are generated in Fig.8.19. The size of the two maximized embedded squares in the butterfly plots represents the magnitude of SNM. In Fig.8.19, (a) represents failure free operation, (b) shows SNM degradation due to TDDB effect, (c) shows the combined effect of TDDB+NBTI on SNM, and (d) is the combination of plots (a) ~ (c) for the sake of easy comparison. These curves are obtained by setting failure circuit model parameters at stress time 0.4 years. It is indicated from these butterfly plots that SRAM cell noise margin shrinks due to TDDB and NBTI stresses, and TDDB has the dominant effect. The gate-to-channel breakdown of  $M3$  leads to symmetrical shrinkage of the two embedded squares, which is distinct

and in contrast to the case of gate-to-diffusion breakdowns presented in [161, 162] where asymmetrical scales of the sizes of the two embedded squares resulted from p-source breakdown. It is expected that gate-to-diffusion breakdown model of TDDB would accelerate SNM degradation. At 0.4 year, even though SNM is significantly reduced, the two transfer curves still cross and form two stable states, therefore, SRAM cell function is maintained.

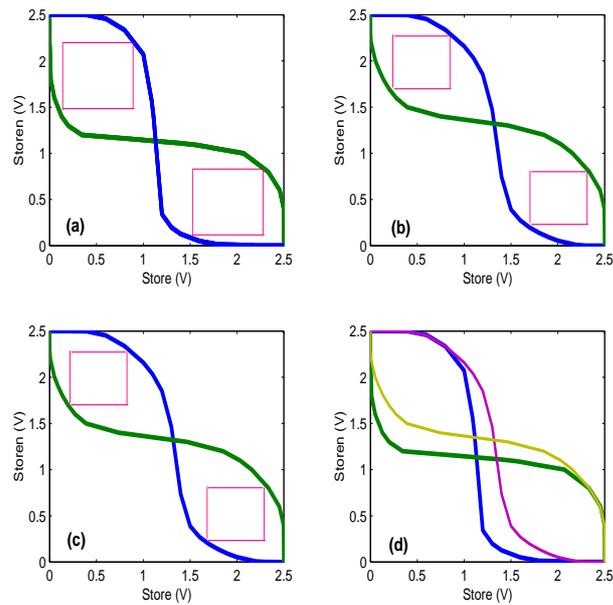


Figure 8.19: Butterfly plots for various failure mechanisms. (a) denotes the no-damage operation, (b) shows SNM degradation due to TDDB, (c) shows the combined effect of TDDB+NBTI, and (d) is the combination of the previous three plots. The difference in (b) and (c) is very small indicating that NBTI is not a dominant effect.

The SRAM circuit survives to 0.4 year but fails at 0.6 year. If the same duty cycle and usage profile are assumed as before, the HCI+TDDB+NBTI lifetime of this SRAM circuit under normal use condition is about 62.5 years.

## 8.6 Reliability Design Techniques

After exploring circuit degradation effects and reliability behaviors with MaCRO models, designers need to make design iterations to improve circuit lifetime if the initial design falls short of reliability specifications. Traditionally, this is an arduous work due to the lack of systematic and convenient reliability analysis method to help pinpoint reliability weak spots and characterize circuit degradation in performance and functionality. With MaCRO models and simulation algorithms, designers can perform a quick reliability analysis and gain knowledge on circuit failure behaviors. Equipped with this reliability knowledge, they can develop their own expertise on reliability improvement through proper design iterations.

In literature, there are some reliability design techniques available for suppressing different failure mechanisms. Reliability design techniques for HCI including transistor sizing, gate topology transform and input signal scheduling are presented in [168]. Some design improvement concept for TDDB is introduced in [169]. A design technique to reduce gate-to-source voltage during static state operation and improve NBTI reliability is introduced in [170]. Even though some progresses have been achieved from individual work, design techniques for TDDB and NBTI have not been thoroughly investigated. With better understanding of circuit reliability behaviors from MaCRO simulation, circuit designer can develop their own guidelines and expertise in this area.

## 8.7 Conclusion

In this chapter, a simple SRAM circuit is designed and simulated with MaCRO models and algorithms to illustrate how to apply this new method to circuit reliability simulation and analysis. Simulation shows HCI and TDDB have significant effects on SRAM cell stability and voltage transfer characteristics, while NBTI mainly degrades cell transition speed when the cell state flips. This case study of SRAM reliability simulation work proves that with MaCRO lifetime models and circuit models, circuit designers will obtain in-depth understanding of circuit failure behaviors and the damage effects of HCI/TDDB/NBTI on circuit operations. Equipped with this knowledge, they can quickly estimate circuit lifetime, make appropriate performance and reliability tradeoffs, and formulate practical design guidelines to improve circuit reliability.

## Chapter 9

### Summary

#### 9.1 Results

Advances in technology have raised many new issues related to both circuit performance and reliability. Today's extremely complex VLSI chips have been designed to gain maximum performance by stretching the limits of available technologies. These limits are often imposed by reliability concerns. As Computer Aided Design (CAD) techniques become more mature and sophisticated, most aspects of modern CMOS VLSI design have been modeled and simulated before committing circuits to silicon. In the CAD tool set, there is obviously a niche for circuit reliability simulation, which will help designers predict device lifetime and failure rate and characterize circuit failure behaviors. Thus designers can make appropriate performance and reliability tradeoffs in the initial design phases.

Even though many advanced DFR tools have been developed, most of them require a large number of simulation iterations and tedious parameter testing work, which limit their real-world applications. One way to effectively overcome these drawbacks is to elevate the focus of reliability analysis from the device wearout process to circuit functionality. Essentially, circuit functional simulation is no more than

solving a group of individual device equivalent circuit model equations to predict the interactions of all these devices upon external stimuli. Therefore, circuit functionality is solely determined by individual device models. From this perspective, circuit degradation or failures can be viewed as the results that device-level wearout effects express themselves at circuit-level by changing their model structures. If the change of device model structures due to wearout effects can be correctly modeled with the inclusion of additional circuit elements, and the relations between these additional elements and the time-dependent wearout parameters can be built and calibrated with simple testing work, then it is foreseeable that circuit reliability simulation will become a natural and simple step of the overall circuit functional simulation.

Starting from this concept, a new SPICE reliability simulation method has been developed, which includes a set of accelerated lifetime models and failure equivalent circuit models for the most common wearout mechanisms including HCI, TDDB and NBTI, respectively. The accelerated lifetime models help to identify the most degraded transistors in a circuit based on their time-variant terminal voltage and current waveforms. Then, failure equivalent circuit models are used to substitute those identified transistors in SPICE simulation to investigate the impact of device wearout on circuit functionality. Device wearout effects are lumped into a limited number of failure equivalent circuit model parameters, and circuit functionality and performance degradation are determined by the magnitude of these model parameters. In this new method, it is unnecessary to perform a large number of iterative SPICE simulation processes. Therefore, simulation time is obviously reduced. Moreover, the model parameters that must be extrapolated have been reduced to

only a small set of failure equivalent circuit elements. So, the reliability testing work becomes less intensive.

This new reliability simulation method can be used for many different application purposes. If all lifetime model parameters are obtained from experimental work, it can accurately calculate circuit lifetime and failure rate. If process parameters of future technologies are projected, it can predict reliability trends over generations and identify critical failure mechanisms. Based on the previous two applications, it can be further used in derating device voltage and temperature for circuit reliability enhancement. If circuit functionality is of primary interest, it can quickly identify more damaged transistors in circuit in terms of the device's terminal voltage and current stress profiles. Then it can include corresponding circuit models in the second-round SPICE simulation, which will reveal whether or not circuit functionality is maintained.

A 17-stage CMOS ring oscillator was simulated with this new method to explore general derating behaviors. It was proven that reduced voltage, frequency and temperature reduces devices's internal stresses, leading to an improvement of device reliability. Since all these variations for a single device are proportional, the ratios can be applied to a full circuit with the help of a simple derating model. Simulation shows that derating factors at different rated voltages nearly follow the same trend. So no matter what the rated voltage is, if voltage is derated to the same ratio, the reliability gain is about the same.

A simple SRAM circuit was designed and simulated to demonstrate how to apply this new method to circuit reliability simulation and analysis. Simulation

showed that for the  $0.25\mu m$  technology, HCI and TDDB have significant effects on SRAM cell stability and VTC while NBTI mainly degrades cell transition speed when the cell state flips.

Properly applying this new method to other typical circuit structures should yield results that either corroborate other researchers' experimental and theoretical work or reveal new phenomena yet to be explored.

## 9.2 Main Contributions

The main contributions of this work in the area of reliability modeling can be summarized as follows: (1) HCI, TDDB and NBTI are treated in a unified framework. Therefore, their relations and interactions can be accounted for in a simple SPICE simulation process. This is not realized in other reliability simulation tools. (2) An existing HCI  $\Delta R_d$  model is improved to include the contribution of both interface trapped charge and oxide trapped charge, the latter one being neglected in the original  $\Delta R_d$  model. This improvement, although complicating parameter extraction work, is physically more comprehensive and accurate in characterizing hot carrier damages. (3) The IBM TDDB lifetime model is improved to cover many important experimental observations of oxide breakdown behaviors, including power law voltage acceleration, non-Arrhenius temperature acceleration, Poisson area scaling statistics and cumulative failure percentile scaling effect. This gives rise to a more comprehensive TDDB lifetime model than the original one. (4) A thorough discussion of available TDDB circuit models is presented, which compensates for

the obvious absence of review papers in this area. (5) Some limitations and an error in the most frequently discussed TDDB circuit model proposed by IMEC are identified. (6) A new TDDB circuit model is proposed and the number of model parameters is reduced to only one, which significantly simplifies TDDB reliability simulation. (7) Weibull statistics is included in explaining the existing IBM NBTI degradation model, which provides a new perspective from which to understand the IBM model and NBTI behaviors. (8) Starting from this IBM NBTI model, a physics and statistics based NBTI lifetime model is developed that is capable of explaining nearly all known NBTI effects, including power law dependence, saturation behaviors and recovery effects. (9) A new NBTI failure equivalent circuit model is developed, which is simple, physics-of-failure based, and expandable. This is the first NBTI failure equivalent circuit model of its kind in literature.

### 9.3 Future Work

As CMOS VLSI technology rapidly advances, this work is far from completion. Even though most of the essential models and algorithms of MaCRO have been developed, we have only taken a first step toward a practical DFR tool. Generating a robust computer program from these models and algorithms is not a trivial work.

At the model level, much work needs to be done, including dynamic stress modeling of each wearout mechanism. Now quasi-static assumption is made to simplify calculation of device operating parameters, which is not accurate and does not reflect device stress profiles in real applications. Another modeling problem is

related to TDDB effects. Currently, there is no TDDB circuit model capable of relating stress time  $t$  to model elements. In this respect, MaCRO is unexceptional. As a result, it is difficult to characterize time evolution of gate oxide breakdown evens at circuit level, and clearly predict when breakdown happens and interferes with circuit operation. Much more experimental and modeling work is required to set the time factor in TDDB circuit models.

Currently, a 6-bit floating-gate high-speed Flash Analog-to-Digital Converter (ADC) is under development. This ADC circuit consists of sample and hold, comparator, and ROM-based digital encoder subcircuits. It is being implemented with a commercial  $0.35\mu m$  technology. Floating-gate MOSFETs explore the potentials of channel carrier injection and oxide tunneling mechanisms to realize charge storage and offset programming functions. However, their special structures and operation configurations arouse many new reliability issues. The reliability behaviors of this floating-gate ADC will be simulated with MaCRO models and algorithms. This will expand MaCRO applications to analog and mixed-signal circuits.

Scientific discovery is the underlying driving force for technology evolution. To make any scientific breakthroughs, research investment must keep pace with technology advancement. As more intellectual and physical resources are devoted to IC reliability simulation, it is foreseeable that DFR will become an indispensable tool sustaining the development of CMOS VLSI technologies.

## Appendix A

### MATLAB Programs for Circuit Model Calculation

#### A.1 Hot Carrier Injection

HCI circuit model parameter calculation in terms of  $0.25\mu m$  technology

*% Parameters obtained from SPICE simulation*

$V_{ds} =$  ;  $V_{gd} =$  ;  $V_{gs} =$  ;

*% Device geometry parameters*

$W =$  ; *% Channel width in  $\mu m$*

$L = 0.25 \times 10^{-6}$ ; *% Channel length in  $\mu m$*

$t =$  ; *% Set the stress time*

*% Physical constants*

$q = 1.6 \times 10^{-19}$ ; *% Electron charge*

$\lambda = 9.2$ ; *% Hot-electron scattering mean-free path in nm*

$\lambda_r = 61.6$ ; *% Re-direction mean-free path in nm*

*% Technology parameters*

$V_t = 0.65$ ; *% Threshold voltage in V*

$t_{ox} = 5.7$ ; *% Oxide thickness in nm*

$\alpha = 2.4 \times 10^{-12}$ ;  $k_n = 2 \times 124.3 \times 10^{-6}$ ;  $C_{ox} = 6.0579 \times 10^{-7}$ ;

*% Current and voltage calculation*

*%  $I_{ds} = k_n \times W/L \times ((V_{gs} - V_t) \times V_{ds} - 1/2 \times V_{ds}^2)$ ; % Linear operation*

*$I_{ds} = 1/2 \times k_n \times W/L \times (V_{gs} - V_t)^2$ ; % Saturation operation*

*$V_{dsat} = (V_{gs} - V_t) \times 12.5 / ((V_{gs} - V_t) + 12.5)$ ;*

*$V_{dff} = V_{ds} - V_{dsat}$ ;*

*% Interface trap calculation*

*$\Delta N_{it} = 2 \times (I_{ds}/W \times \exp(-5.5224/V_{dff}) \times t)^{0.65} \times 10^4$ ;*

*% Oxide trap calculation*

*$E_{ox} = V_{gd}/t_{ox}$ ;  $\lambda E_m = \lambda \times V_{dff}/10$ ;*

*$\Phi_b = 3.2 - 2.59 \times 10^{-4} \times (10^7)^{1/2} \times (E_{ox})^{1/2} - 4 \times 10^{-5} \times (10^7)^{2/3} \times (E_{ox})^{2/3}$ ;*

*$I_{ei} = 1/2 \times t_{ox}/\lambda_r \times (\lambda E_m/\Phi_b)^2 \times \exp(-\Phi_b/\lambda E_m) \times I_{ds}$ ;*

*$\Delta N_{ox} = 1.58 \times 10^{12} \times (1 - e^{(-4.5 \times 10^{-4} \times I_{ei} \times t)}) - 1.36 \times 10^{12} \times (1 - e^{(-1 \times 10^{-4} \times I_{ei} \times t)})$ ;*

*%  $\Delta R_d$  calculation*

*$\Delta N = \Delta N_{it} + \Delta N_{ox}$ ;  $V_{gdx} = V_{gd} - 0.65$ ;*

*$V_{Rd} = -V_{gdx} + \sqrt{(V_{gdx})^2 + 2V_{ds}(\alpha \Delta N (V_{gdx} + 0.5V_{ds}) / (1 + \alpha \Delta N) + q \Delta N / C_{ox})}$ ;*

*$\Delta R_d = (1 + \alpha \times \Delta N) / I_{ds} \times V_{Rd}$*

## A.2 Time Dependent Dielectric Breakdown

TDDDB circuit model parameter calculation in terms of  $0.25\mu m$  technology

*% Parameters from SPICE simulation*

*$V_{g=} ; V_{d=} ; V_{s=} ;$*

*% Device geometry parameters*

$W =$  ; *% Channel width in  $\mu\text{m}$*

$L = 0.25 \times 10^{-6}$ ; *% Channel length in  $\mu\text{m}$*

$L_1 = L/2$ ; *% Breakdown location*

*% Process parameters*

$t_{ox} = 5.7$ ; *% Oxide thickness in nm*

$C_{ox} = 6.0579 \times 10^{-7}$ ;

$\Phi_b = 0.026 \times \log(N_{ch}/(1.5 \times 10^{10}))$ ;

*% nMOS parameters*

*%  $V_t = 0.45$ ; %  $k = 2 \times 124.3 \times 10^{-6}$ ; %  $N_{ch} = 2.3549 \times 10^{17}$ ; %  $\gamma = 0.43$ ;*

*% pMOS parameters*

$V_t = -0.54$ ;  $k = -2 \times 24.5 \times 10^{-6}$ ;  $N_{ch} = 4.1589 \times 10^{17}$ ;  $\gamma = -0.61$ ;

*% Model parameter calculation*

$I_{ds} = k \times W/L \times ((V_g - V_s - V_t) \times (V_d - V_s) - 1/2 \times (V_d - V_s)^2)$ ; *% Linear operation*

*%  $I_{ds} = 1/2 \times k \times W/L \times (V_g - V_s - V_t)^2$ ; % Saturation operation*

$V_i = (V_g - V_t) - \sqrt{(V_g - V_t)^2 - (V_s^2 + 2(V_g - V_s - V_t) \times V_s + 2I_{ds}L_1/(kW))}$ ;

*%  $V_{tb} = V_t + \gamma \times (\sqrt{|2 \times \Phi_b + V_i|} - \sqrt{|2 \times \Phi_b|})$ ; % nMOS*

$V_{tb} = V_t + \gamma \times (\sqrt{|-2 \times \Phi_b - V_i|} - \sqrt{|-2 \times \Phi_b|})$ ; *% pMOS*

$I_d = k \times W/(L - L_1) \times ((V_g - V_i - V_{tb}) \times (V_d - V_i) - 1/2 \times (V_d - V_i)^2)$

$I_s = k \times W/L_1 \times ((V_g - V_s - V_t) \times (V_i - V_s) - 1/2 \times (V_i - V_s)^2)$

$I_{ox} = I_s - I_d$

### A.3 Negative Bias Temperature Instability

NBTI circuit model parameter calculation in terms of  $0.25\mu m$  technology

*% Parameters from SPICE simulation*

$V_{gd} = ; V_{gs} = ;$

*% Physical constants*

$q = 1.6 \times 10^{-19}; \kappa = 8.62 \times 10^{-5};$  *% Boltzmann's constant in eV/Kelvin*

$t = ;$  *% Stress time*

*% Process parameters*

$t_{ox} = 5.7 \times 10^{-7};$  *% Oxide thickness in cm*

$t_{oxn} = t_{ox} \times 10^7;$  *% Oxide thickness in nm*

$k_{ox} = 3.5 \times 10^{-13};$  *% SiO<sub>2</sub> permittivity in F/cm<sup>2</sup>*

*% Model fitting parameters*

$\tau = 8 \times 10^7; \beta = 0.3;$

$n = 5; K = 3 \times 10^{-6};$  *% Leakage model parameter*

$H_0 = 1.5 \times 10^{12};$  *% in unit /cm<sup>2</sup>*

$T = 300;$  *% Temperature in Kelvin*

$E_{fx} = -0.16; E_f = 0.98; E_1 = 0.1;$  *% in unit eV*

$E_2 = E_{fx} - E_f + 5.9412 \times ((V_{gs} - 0.2)/t_{oxn})^{2/3};$  *% in unit eV*

*% Model elements calculation*

$\Delta V_{max} = q \times H_0 \times (1/(1 + 2e^{-E_1/\kappa T}) + 1/(1 + 2e^{-E_2/\kappa T})) \times t_{ox}/k_{ox};$

$R_G = \Delta V_{max}/(K \times V_{gd}^n + K \times V_{gs}^n) \times (1 - \exp(-(t/\tau)^\beta))$

$I_{GS} = K \times V_{gs}^n; I_{GD} = K \times V_{gd}^n$

# Appendix B

## Flowchart of Lifetime and Reliability Trend Prediction

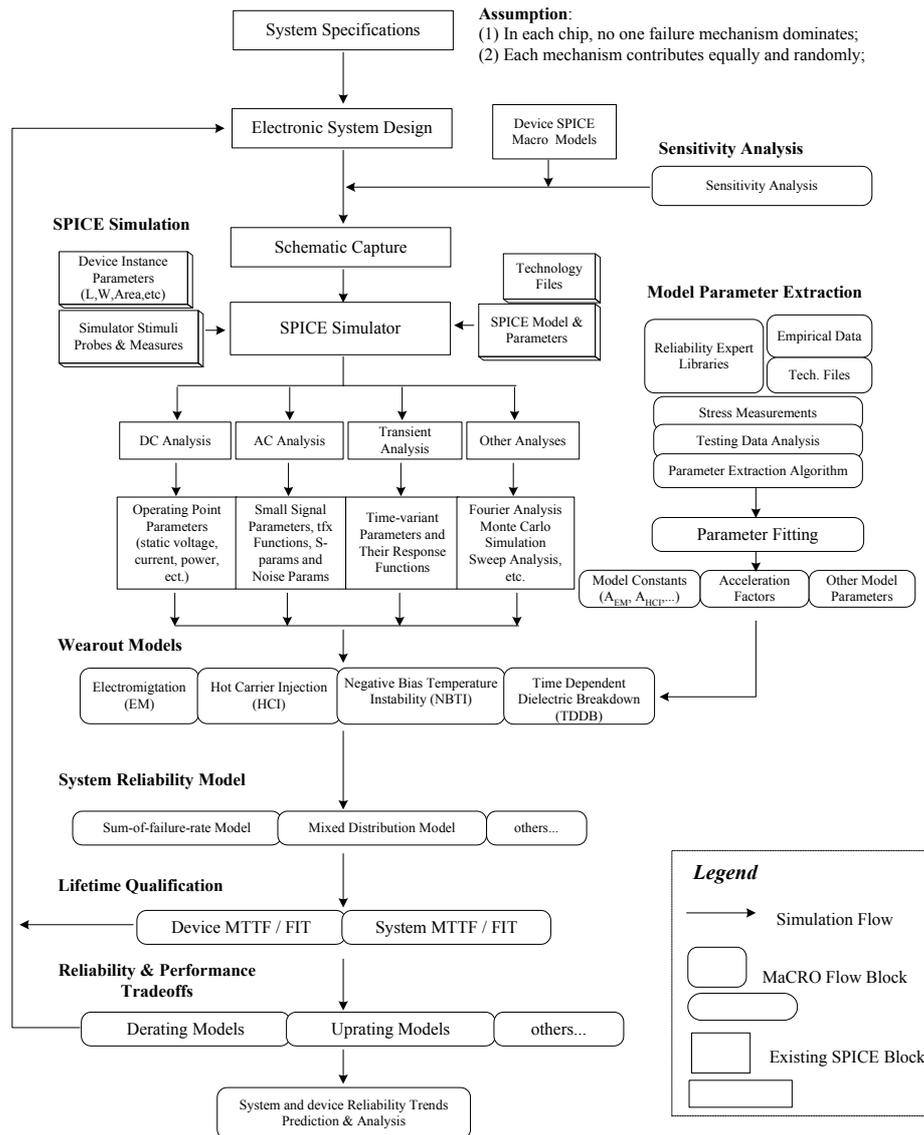


Figure B.1: MaCRO Flow of lifetime, failure rate and reliability trend prediction.

## Appendix C

### Terms and Abbreviations

ADC	Analog-to-Digital Converter
AHI	Anode Hole Injection
ALT	Accelerated Life Testing
AST	Accelerated Stress Tests
ATPG	Automatic Test Pattern Generation
BERT	Berkeley Reliability Tools
BIR	Built-In-Reliability
BTI	Biased Temperature Instability
CAD	Computer Aided Design
CHC	Channel Hot Carrier
COTS	Commercial-Off-The-Shelf
DFR	Design For Reliability
EM	Electromigration
FIT	Failure in time
FN	Fowler-Nordheim Tunneling
GCA	Gradual Channel Approximation
GIDL	Gate-Induced Drain Leakage
GOS	Gate Oxide Short

HCI	Hot Carrier Injection
HISREM	Hot Carrier Induced Series Resistance Enhancement Model
ITRS	International Technology Roadmap for Semiconductor
KCL	Kirchhoff's Current Law
LNA	Low Noise Amplifier
MaCRO	Maryland Circuit Reliability-Oriented SPICE simulation method
MSM	Matrix Stressing Method
MTTF	Mean Time To Failure
NBTI	Negative Bias Temperature Instability
PBTI	Positive Bias Temperature Instability
RAMP	Reliability Aware Micro-Processor
SNM	Static Noise Margin
SoC	System-on-Chip
SOFR	Sum-of-failure-rates
S-L	Shatzkes and Liloyd model
TCAD	Technology Computer Aided Design
TDDB	Time Dependent Dielectric Breakdown
VTC	Voltage Transfer Characteristics

## BIBLIOGRAPHY

- [1] R. Blish, T. Dellin, S. Huber, et al., "Critical Reliability Challenges for The International Technology Roadmap for Semiconductor (ITRS)," Technology Transfer #03024377A-TR, International SEMATECH 2003.
- [2] "International Technology Roadmap for Semiconductors," *ITRS 2003 Edition*.  
<http://public.itrs.net/>
- [3] Technical Brief, "Calculating MTTF When You Have Zero Failures," Relex Software Corporation.
- [4] C. Hu, "Future CMOS scaling and reliability," *Proceedings of the IEEE*, vol.81, No.5, pp.682-689, May 1993.
- [5] S. Aur, D. E. Hocevar, and P. Yang, "HOTRON - A circuit hot electron effect simulator," *IEEE International Conference on Computer-Aided Design, ICCAD'87, Digest of Technical Papers*, pp.256-259, Nov. 1987.
- [6] B. J. Sheu, W. J. Hsu, and B. W. Lee, "An integrated-circuit reliability simulator-RELY," *IEEE Journal of Solid-State Circuits*, vol.24, no.2, pp.473-477, Apr. 1989.
- [7] A. T. Yang, and S. M. Kang, "iSMILE: a novel circuit simulation program with emphasis on new device model development," *Proceedings of the 26th ACM/IEEE conference on Design automation, DAC'89*, pp.630-633, Jun. 1989.

- [8] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami, K. Quader, P. K. Ko, and C. Hu, "Berkeley reliability tools-BERT," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol.12, no.10, pp.1524-1534, Oct. 1993.
- [9] X. D. Xuan, A. Chatterjee, A. D. Singh, N. P. Kim, and M. T. Chisa, "IC reliability simulator ARET and its application in design-for-reliability," *Proceedings of the 12th Asian Test Symposium, ATS'03*, pp.18-21, Nov. 2003.
- [10] Y. H. Shih, Y. Leblebici, and S. M. Kang, "ILLIADS: a fast timing and reliability simulator for digital MOS circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol.12, no.9, pp.1387-1402, Sept. 1993.
- [11] S. S. Chung, T. S. Chang, and P.C. Hsu, "A high level simulator feasible for reliability analysis of VLSI circuits," *Proceedings Fourth Annual IEEE International ASIC Conference and Exhibit*, pp.15-2/1-4, Sept. 1991.
- [12] L. F. Wu, J. Fang, H. Yonezawa, and Y. Kawakami, "GLACIER: a hot carrier gate level circuit characterization and simulation system for VLSI design," *Proceedings IEEE 2000 First International Symposium on Quality Electronic Design, ISQED'00*, pp.73-79, Mar. 2000.
- [13] Y. Leblebici, S.M. Kang, "Hot-Carrier Reliability of MOS VLSI Circuits," pp. 61, Kluwer Academic Publishers, Norwell, MA, 1993. ISBN 0-7923-9352-X.

- [14] S. Minehane, R. Duane, P. O'Sullivan, K. G. McCarthy, A. Mathewson, "Design for Reliability," *Microelectronics Reliability*, vol.40, pp.1285-1294, 2000.
- [15] "Cadence Virtuoso UltraSim Full Chip Simulator Datasheet," Cadence Design Systems, <http://www.cadence.com/>
- [16] C. Hu, "IC Reliability Simulation," *IEEE Journal of Solid-State Circuits*, vol.27, no.3, pp.241-246, Mar. 1992.
- [17] Cadence White paper, "Reliability Simulation in Integrated Circuit Design," <http://www.cadence.com/>.
- [18] K. N. Quader, C. C. Li, R. Tu, et al., "A Bidirectional NMOSFET Current Reduction Model for Simulation of Hot-Carrier-Induced Circuit Degradation," *IEEE Transactions on Electron Devices*, vol.40, no.12, pp.2245-2254, Dec. 1993.
- [19] M. Karam, W. Fikry, and H. Ragai, "Implementation of Hot-Carrier Reliability Simulation in Eldo," *Mentor Graphics Deep Submicron Technical Publication*, Sep. 2000.
- [20] P. C. Li, and I. N. Hajj, "Computer-aided redesign of VLSI circuits for hot-carrier reliability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol.15, no.5, pp.453-464, May 1996.
- [21] W. Jiang, H. Le, J. Chung, and T. Kopley, "Assessing circuit-level hot-carrier reliability," *IEEE Proceedings of 36th Annual International Reliability Physics Symposium*, IRPS'98, pp.173-179, Mar. 1998.

- [22] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami, K. Quader, P. K. Ko, and C. Hu, "Berkeley reliability tools-BERT," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol.12, no.10, pp.1524-1534, Oct. 1993.
- [23] G. La Rosa, F. Guarin, S. Rauch, A. Acovic, J. Lukaitis, and E. Crabbe, "NBTI-channel hot carrier effects in PMOSFETs in advanced CMOS technologies," *IEEE Proceedings of 35th Annual International Reliability Physics Symposium, IRPS'97*, pp.282-286, Apr. 1997.
- [24] C. Z. Yu, Y. Liu, A. Sadat, and J. S. Yuan, "Impact of temperature-accelerated voltage stress on PMOS RF performance," *IEEE Transactions on Device and Materials Reliability*, vol.4, no.4, pp.664-669, Dec. 2004.
- [25] K. Sakui, S. S. Wong, and B. A. Wooley, "The Effects Of Hot Carriers Generation On The Operation Of Neighboring Devices And Circuits," *1993 Symposium on VLSI Technology*, Digest of Technical Papers, pp.11-12, May. 1993.
- [26] T. Horiuchi, "In-circuit hot-carrier model and its application to inverter chain optimization," *IEEE Transactions on Electron Devices*, vol.43, no.9, pp.1428-1432, Sept. 1996.
- [27] Xiaojun Li, B. Huang, J. Qin, X. Zhang, M. Talmor, Z. Gur, and J. B. Bernstein, "Deep submicron CMOS integrated circuit reliability simulation with SPICE," *IEEE Sixth International Symposium on Quality of Electronic Design, ISQED'05*. pp.382-389, Mar. 2005.

- [28] M. M. Kuo, K. Seki, P. M. Lee, J. Y. Choi, P. K. Ko, and C. Hu, "Simulation of MOSFET Lifetime under AC Hot-Electron Stress," *IEEE Transactions on Electron Devices*, vol.35, no.7, pp.1004-1011, Jul. 1988.
- [29] W. J. Hsu, B. J. Sheu, S. M. Gowda, and C. G. Hwang, "Advanced Integrated-Circuit Reliability Simulation Including Dynamic Stress Effects," *IEEE Journal of Solid-State Circuits*, vol.27, no.3, pp.247-257, Mar. 1992.
- [30] Y. Leblebici, S. M. Kang, "Simulation of Hot-Carrier Induced MOS Circuit Degradation for VLSI Reliability Analysis," *IEEE Transactions on Reliability*, vol.43, no.2, pp.197-206, Jun. 1994.
- [31] N. Hwang, L. Forbes, "Hot-Carrier Induced Series Resistance Enhancement Model (HISREM) of nMOSTFET's for Circuit Simulation and Reliability Projections," *Microelectron. Reliab.*, vol.35, issue 2, pp.225-239, 1995.
- [32] E. Y. Wu, E. J. Nowak, A. Vayshenker, W. L. Lai, and D. L. Harmon, "CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics," *IBM J. RES & DEV*, vol.46, no.2/3, Mar./May 2002.
- [33] E. Wu, J. Sune, W. Lai, etc., "Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin oxides," *Microelectronic Engineering*, vol.59, pp.25-31, 2001.
- [34] E. Wu, J. Sune, W. Lai, E. Nowark, L. McKenna, A. Vayshenker, and D. Harmon, "Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin gate oxides," *Solid-State Electronics*, vol.46, pp.1787-1798, 2002.

- [35] E. Wu, and J. Sune, "Power-Law Voltage Acceleration: A Key Element for Ultra-thin Gate OXide Reliability," *Microelectronics Reliability*, Invited paper in special issue of *Microelectronics Reliability*, 2005.
- [36] S. Zafar, B. Lee, J. Stathis, A. Callegari, and T. Ning, "A Model for Negative Bias Temperature Instability (NBTI) in Oxide and High  $\kappa$  pFETs," *IEEE Symposium on VLSI Technology*, Digest of Technical Papers, pp.208-209, 2004.
- [37] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation," *Journal of Applied Physics*, vol.97, issue 1, pp.1-9, 2005.
- [38] JESD85. "Methods for Calculating Failure Rates in Units of FITs," *JEDEC Publication*, JEDEC Solid State Technology Association, July 2001.
- [39] J. Srinivasan, S. V. Adve, P. Bose, J. A. Rivers, and C.-K. Hu, "RAMP: A Model for Reliability Aware MicroProcessor Design," *IBM Research Report*, RC23048, Dec. 2003.
- [40] JEP122B, "Failure Mechanisms and Models for Semiconductor Devices," JEDEC Publication, JEDEC Solid State Technology Association, Aug. 2003.
- [41] P. C. Hsu, S. Tam, "Relationship Between MOSFET Degradation and Hot-Electron-Induced Interface-State Generation," *IEEE Electron Device Letters*, vol.EDL-5, no.2, pp.50-52, Feb. 1984.
- [42] D. K. Slisher, R. G. Filippi, D. W. Storaska, et al., "Scaling of Si MOSFETs For Digital Applications," Rensselaer Polytechnic Institute, Dec. 1999.

- [43] G. Groeseneken, R. Bellens, G. Van den bosch, et al., "Hot-carrier degradation in submicrometre MOSFETs: from uniform injection towards the real operating conditions," *Semicon. Sci. Technol.*, vol.10, pp.1208-1220, 1995.
- [44] G. Groeseneken, R. Degraeve, T. Nigam, et al., "Hot carrier degradation and time-dependent dielectric breakdown in oxides," *Microelectronic Engineering*, vol.49, pp.27-40, 1999.
- [45] 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, no.1, pp.295-305, Feb. 1985.
- [46] Y. Leblebici, S.M. Kang, "Hot-Carrier Reliability of MOS VLSI Circuits," pp. 61, Kluwer Academic Publishers, Norwell, MA, 1993. ISBN 0-7923-9352-X.
- [47] Y. Taur, T.H. Ning, "Fundamentals of modern VLSI devices," pp. 157, Cambridge University Press, 1998. ISBN 0-521-55959-6.
- [48] R. H. Tu, E. Rosenbaum, W. Y. Chan, C. C. Li, E. Minami, K. Quader, P. K. Ko, and C. Hu, "Berkeley reliability tools-BERT," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol.12, no.10, pp.1524-1534, Oct. 1993.
- [49] W. J. Hsu, B. J. Sheu, S. M. Gowda, and C. G. Hwang, "Advanced Integrated-Circuit Reliability Simulation Including Dynamic Stress Effects," *IEEE Journal of Solid-State Circuits*, vol.27, no.3, pp.247-257, Mar. 1992.

- [50] W. Liu, X. Jin, J. Chen, M. Jeng, Z. Liu, Y. Cheng, K. Chen, M. Chan, K. Hui, J. Huang, R. Tu, P. K. Ko, and C. Hu, "BSIM3v3.2.2 MOSFET Model Users' Manual," University of California, Berkeley, 1999.
- [51] P. C. Li, G. I. Stamoulis, and I. N. Hajj, "iProbe-d: a Hot-Carrier and Oxide Reliability Simulator," *IEEE Proceedings of International Reliability Physics Symposium*, IRPS'94, pp.274-279, 1994.
- [52] K. N. Quader, C. C. Li, R. Tu, E. Rosenbaum, P. K. Ko, and C. Hu, "A Bidirectional NMOSFET Current Reduction Model for Simulation of Hot-Carrier-Induced Circuit Degradation," *IEEE Transactions on Electron Devices*, vol.40, no.12, pp.2245-2254, Dec. 1993.
- [53] K. N. Quader, C. Li, R. Tu, E. Rosenbaum, P. Ko, and C. Hu, "A New Approach for Simulation of Circuit Degradation Due to Hot-Electron Damage in NMOSFETs," *IEEE International Electron Devices Meeting 1991*, IEDM'91 Technical Digest, pp.337-340, Dec. 1991.
- [54] Y. Leblebici, S. M. Kang, "Modeling of nMOS Transistors for Simulation of Hot-Carrier-Induced Device and Circuit Degradation," *IEEE Transactions on Computer-Aided Design*, vol.11, No.2, pp.235-246, Feb. 1992.
- [55] Y. Leblebici, S. M. Kang, "Modeling and Simulation of Hot-Carrier-Induced Device Degradation in MOS circuits," *IEEE Journal of Solid-State Circuits*, vol.28, No.5, pp.585-595, May 1993.

- [56] N. Hwang, L. Forbes, "Hot-Carrier Induced Series Resistance Enhancement Model (HISREM) of nMOSTFET's for Circuit Simulation and Reliability Projections," *Microelectron. Reliab.*, vol.35, issue 2, pp.225-239, 1995.
- [57] C. C. Li, K. N. Quader, E. R. Minami, C. Hu, and P. K. Ko, "A New Bidirectional PMOSFET Hot-Carrier Degradation Model for Circuit Reliability Simulation," *IEEE International Electron Devices Meeting 1992, IEDM'92*, pp.547-550, 1992.
- [58] "International Technology Roadmap for Semiconductors - 2003 Edition," ITRS'03, <http://public.itrs.net/>
- [59] R. Degraeve, B. Kaczer, G. Groeseneken, "Reliability: a possible showstopper for oxide thickness scaling," *Semicon. Sci. Technol.*, vol.15, pp.436-44, 2000.
- [60] J. S. Suehle, "Ultrathin Gate Oxide Reliability: Physical Models, Statistics, and Characterization," *IEEE Transactions on Electron Devices*, vol.49, no.6, pp.958-971, Jun. 2002.
- [61] Y. Chen, J. S. Suehle, C. C. Shen, J. Bernstein, C. Messick, and P. Chappala, "The Correlation of Highly Accelerated  $Q_{bd}$  Tests to TDDB Life Tests for Ultra-Thin Gate Oxides," *IEEE Annual International Reliability Physics Symposium, IRPS'98*, pp.87-91, 1998.
- [62] JEP122B, "Failure Mechanisms and Models for Semiconductor Devices," JEDEC Publication, JEDEC Solid State Technology Association, Aug. 2003.

- [63] Y. C. Yeo, Q. Lu, and C. Hu, "MOSFET Gate Oxide Reliability: Anode Hole Injection Model and its Applications," *International Journal of High Speed Electronics and Systems*, vol.11, no.3, pp.849-886, 2001.
- [64] N. Shiono, M. Itsumi, "A Lifetime Projection Method Using Series Model and Acceleration Factors for TDDB failures of Thin Gate Oxides," *IEEE Annual International Reliability Physics Symposium*, IRPS'93, pp.1-6, 1993.
- [65] C. Hu, Q. Lu, "A Unified Gate Oxide Reliability Model," *IEEE Annual International Reliability Physics Symposium*, IRPS'99, pp.47-51, 1999.
- [66] G. Groeseneken, R. Degraeve, T. Nigam, et al., "Hot carrier degradation and time-dependent dielectric breakdown in oxides," *Microelectronic Engineering*, vol.49, pp.27-40, 1999.
- [67] F. Monsieur, E. Vincent, D. Roy, S. Bruyere, G. Pananakakis, and G. Ghibaudo, "Time to Breakdown and Voltage to Breakdown modeling for Ultra0thin Oxides ( $T_{ox} < 32\text{\AA}$ )," *2001 IRW Final Report*, pp.20-25, 2001.
- [68] G. Ribes, S. Bruyere, F. Monsieur, D. Roy, and V. Huard, "New insights into the change of voltage acceleration and temperature activation of oxide breakdown," *Microelectronics Reliability*, vol.43, issue 8, pp.1211-1214, Aug. 2003.
- [69] F. Monsieur, E. Vincent, D. Roy, S. Bruyere, J. C. Vildeuil, G. Pananakakis, and G. Ghibaudo, "A thorough Investigation of Progressive Breakdown in

Ultra-thin Oxides. Physical Understanding and Application for Industrial Reliability Assessment,” *IEEE Annual International Reliability Physics Symposium*, IRPS’02, pp.45-54, 2002.

- [70] E. Y. Wu, E. J. Nowak, A. Vayshenker, W. L. Lai, and D. L. Harmon, “CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics,” *IBM J. RES & DEV*, vol.46, no.2/3, Mar./May 2002.
- [71] E. Wu, J. Sune, W. Lai, etc., “Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin oxides,” *Microelectronic Engineering*, vol.59, pp.25-31, 2001.
- [72] E. Wu, J. Sune, W. Lai, E. Nowark, L. McKenna, A. Vayshenker, and D. Harmon, “Interplay of voltage and temperature acceleration of oxide breakdown for ultra-thin gate oxides,” *Solid-State Electronics*, vol.46, pp.1787-1798, 2002.
- [73] E. Wu, and J. Sune, “Power-Law Voltage Acceleration: A Key Element for Ultra-thin Gate OXide Reliability,” *Microelectronics Reliability*, Invited paper in special issue of Microelectronics Reliability, 2005.
- [74] J. Srinivasan, S. V. Adve, P. Bose, J. A. Rivers, and C.-K. Hu, “RAMP: A Model for Reliability Aware MicroProcessor Design,” *IBM Research Report*, RC23048, Dec. 2003.
- [75] J. Srinivasan, S, V, Adve, P. Bose, and Jude A. Rivers, “The Case for Lifetime Reliability-Aware Microprocessor,” *IEEE Proceedings of 31<sup>st</sup> Annual International Symposium on Computer Architecture*, ISCA’04, Jun. 2004.

- [76] 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, no.1, Mar. 2001.
- [77] J. H. Stathis, "Physical Models of Ultra Thin Oxide Reliability in CMOS Devices and Implications for Circuit Reliability," *IWGI 2001*, pp.26-29, 2001.
- [78] J. H. Stathis, R. Rodriguez, and B. P. Linder, "Circuit implications of gate oxide breakdown," *Microelectronics Reliability*, vol.43, pp.1193-1197, 2003.
- [79] R. Rodriguez, J. H. Stathis, B. P. Linder, etc., "The Impact of Gate-Oxide Breakdown on SRAM Stability," *IEEE Electron Device Letters*, vol.23, no.9, pp.559-561, Sep. 2002.
- [80] R. Rodriguez, J. H. Stathis, and B. P. Linder, "A Model for Gate-Oxide Breakdown in CMOS Inverters," *IEEE Electron Device Letters*, vol.24, no.2, pp.114-116, Feb. 2003.
- [81] J. Segura, C. D. Benito, A. Rubio, and C. F. Hawkins, "A Detailed Analysis of GOS Defects in MOS Transistors: Testing Implications at Circuit Level," *IEEE International Test Conference*, pp.544-551, 1995.
- [82] J. Segura, C. D. Benito, A. Rubio, and C. F. Hawkins, "A Detailed Analysis and Electrical Modeling of Gate Oxide Shorts in MOS Transistors," *Journal of Electronic Testing: Theory and Applications*, vol.8, pp.229-239, 1996.

- [83] A. Avellan, W. H. Krautschneider, "Impact of Soft and Hard Breakdown on Analog and Digital Circuits," *IEEE Transactions on Device and Materials Reliability*, vol.4, no.4, pp.676-680, 2004.
- [84] H. Yang, J. S. Yuan, Y. Liu, and E. Xiao, "Effect of Gate Oxide Breakdown on RF Performance," *IEEE Transactions on Device and Materials Reliability*, vol.3, no.3, pp.93-97, Sep. 2003.
- [85] H. Yang, J. S. Yuan, and E. Xiao, "Effect of Gate Oxide Breakdown on RF Device and Circuit Performance," *IEEE Annual International Reliability Physics Symposium*, IRPS'03, pp.1-4, 2003.
- [86] B. Kaczer, R. Degraeve, M. Rasras, A. De Keersgieter, K. Van De Mieroop, and G. Groeseneken, "Analysis and modeling of a digital CMOS circuit operation and reliability after gate oxide breakdown: a case study," *Microelectronics Reliability*, vol.42, pp.555-564, 2002.
- [87] B. Kaczer, R. Degraeve, M. Rasras, K. Van De Mieroop, P. J. Roussel, and G. Groeseneken, "Impact of MOSFET Gate Oxide Breakdown on Digital Circuit Operation and Reliability," *IEEE Transactions on Electron Devices*, vol.49, no.3, pp.500-506, Mar. 2002.
- [88] B. Kaczer, R. Degraeve, A. De Keersgieter, K. Van De Mieroop, V. Simons, and G. Groeseneken, "Consistent Model for Short-Channel nMOSFET After Hard Gate Oxide Breakdown," *IEEE Transactions on Electron Devices*, vol.49, no.3, pp.507-513, Mar. 2002.

- [89] R. Degraeve, B. Kaczer, A. De Keersgieter, and G. Groeseneken, "Relation Between Breakdown Mode and Location in Short-Channel nMOSFETs and Its Impact on Reliability Specifications," *IEEE Transactions on Device and Materials Reliability*, vol.1, no.3, pp.163-169, Sep. 2001.
- [90] B. Kaczer, F. Crupi, R. Degraeve, M. Rasras, Ph. Roussel, C. Ciofi, and G. Groeseneken, "Observation of hot-carrier-induced nFET gate-oxide breakdown in dynamically stressed CMOS circuits," *IEEE IEDM 1992*, pp.171-174, 2002.
- [91] B. Kaczer, R. Degraeve, A. De Keersgieter, K. Van De Mieroop, T. Bearda, and G. Groeseneken, "Consistent model for short-channel nMOSFET post-hard-breakdown characteristics," *Symposium on VLSI Technology*, Digest of Technical Papers, pp.121-122, 2001.
- [92] M. Shaheen, S. Mourad, "Gate to Channel Shorts in pMOS Devices: Effects on Logic Gate Failures," *Proceedings of IEEE International Symposium on Circuits and Systems, ISCAS'98*, vol.6, pp.211-214, 1998.
- [93] X. Lu, Z. Li, W. Qiu, H. Walker and W. Shi, "A Circuit Level Fault Model for Resistive Shorts of MOS Gate Oxide," *5th International Workshop on Microprocessor Test and Verification*, Sep. 2004.
- [94] T.-S. Yeoh, N. R. Kamat, R. S. Nair, and S.-J. Hu, "Gate oxide breakdown model in MOS transistors," *Proceedings of IEEE International Reliability Physics*, pp.149-155. 1995.

- [95] T.-S. Yeoh, and S.-J. Hu, "Influence of MOS transistor gate oxide breakdown on circuit performance," *Proceedings IEEE International Conference on Semiconductor Electronics*, pp.59-63. 1998.
- [96] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "MDelay Testing of MOS Transistor with Gate Oxide Short," *Proceedings of the Twelfth Asian Symposium, ATS'03*, pp.168-173, 2003.
- [97] M. Renovell, J. M. Galliere, F. Azais, and Y. Bertrand, "Modeling gate oxide short defects in CMOS minimum transistors," *Proceedings of the Seventh IEEE European Test Workshop*, pp.15-20, 2002.
- [98] D. Schroder, J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *Journal of Applied Physics*, vol.94, issue 1, pp.1-18, Jul. 2003.
- [99] M. A. Alam, and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronics Reliability*, vol.45, pp.71-81, 2005.
- [100] S. Zafar, B. Lee, J. Stathis, A. Callegari, and T. Ning, "A Model for Negative Bias Temperature Instability (NBTI) in Oxide and High  $\kappa$  pFETs," *IEEE Symposium on VLSI Technology, Digest of Technical Papers*, pp.208-209, 2004.
- [101] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation," *Journal of Applied Physics*, vol.97, issue 1, pp.1-9, 2005.

- [102] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, "A thorough investigation of MOSFETs NBTI degradation," *Microelectronics Reliability*, vol.45, pp.83-98, 2005.
- [103] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectronics Reliability*, Introductory Invited Paper, Article in press, 2005.
- [104] V. Reddy, J. Carulli, A. Krishnan, W. Bosch, and B. Burgess, "Impact of Negative Bias Temperature Instability on Production Parameter Drift," *IEEE ITC International Test Conference*, pp.148-155, 2004.
- [105] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost and S. Krishnan, "Impact of negative bias temperature instability on digital circuit reliability," *Microelectronics Reliability*, vol.45, pp.31-38, 2005.
- [106] A. S. Goda, and G. Kapila, "Design for degradation CAD tools for managing transistor degradation mechanisms", *IEEE Proceedings of the sixth International Symposium on Quality Electronic Design, ISQED'05*, Section 4B, pp.416-420, Mar. 2005.
- [107] S. Chakravarthi, A. T. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, "A Comprehensive Framework For Predictive Modeling of Negative Bias Temperature Instability," *Proceedings of Annual IEEE International Reliability Physics*, pp.273-282. 2004.

- [108] 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.
- [109] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI saturation effect and its impact on electric field dependence of the lifetime," *Microelectronics Reliability*, Article in press, 2005.
- [110] A. Narr, and A. Lill, "Lifetime Prediction for PMOS and NMOS Devices Biased on A Degradation Model for Gate-Bias-Stress," *Microelectronics Reliability*, vol.37, no. 10/11, pp.1433-1436, 1997.
- [111] Y. H. Lee, S. Jacobs, S. Stadler, N. Mielke, and R. Nachman, "The impact of PMOST bias-temperature degradation on logic circuit reliability for performance," *Microelectronics Reliability*, vol.45, pp.107-114, 2005.
- [112] X. J. Li, B. Huang, J. Qin, X. Zhang, M. Talmor, Z. Gur, and J. B. Bernstein, "Deep Submicron CMOS Integrated Circuit Reliability Simulation with SPICE," *IEEE Proceedings of the Sixth International Symposium on Quality Electronic Design*, ISQED'05, Section 4A, pp.382-389, 2005.
- [113] G. La Rosa, "NBTI Challenges in PMOSFETs of Advanced CMOS Technologies," *IEEE International Reliability Physics Symposium*, IRPS'03 Tutorial Notes, Mar. 2003.

- [114] R. Thewes, R. Brederlow, C. Schlunder, P. Wiczorek, B. Ankele, A. Hesener, L. Holz, S. Kessek, and W. Weber, "MOS Transistor Reliability under Analog Operation", *Microelectronics Reliability*, vol.40, pp.1545-1554, 2000.
- [115] Y. F. Chen, M. H. Lin, C. H. Chou, et al., "Negative bias temperature instability (NBTI) in deep sub-micron p<sup>+</sup>-gate pMOSFETs," *IEEE Integrated Reliability Workshop Final Report*, pp.98-101, 2000.
- [116] N. T. Do, T. Q. Vu, G. Warren, et al., "Negative bias instability in silicon-on-sapphire n-channel MOSFETs". *Proceedings IEEE International SOI Conference*, pp.85-86, 1998.
- [117] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, "Dynamic NBTI of p-MOS Transistors and Its Impact on MOSFET Scaling,". *IEEE Electron Device Letters*, vol.23, no.12, pp.734-736, Dec. 2002.
- [118] J. Srinivasan, S. V. Adve, P. Bose and J. Rivers, "Exploring Structural Duplication for Lifetime Reliability Enhancement,". *Proceedings of the 32nd International Symposium on Computer Architecture*, ISCA'05, Jun. 2005.
- [119] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost and S. Krishnan, "Impact of negative bias temperature instability on digital circuit reliability," *IEEE Annual International Reliability Physics Symposium*, pp.248-254, 2002.

- [120] H. Puchner, L. Hinh, “NBTI reliability analysis for a 90 nm CMOS technology,” *Proceeding of the 34th European Solid-State Device Research conference, ESSDERC’04*, pp.257-260, 2004.
- [121] W. Abadeer, and W. Ellis, “Behavior of NBTI Under AC Dynamic Circuit Conditions,” *IEEE Annual International Reliability Physics Symposium*, pp.17-22, 2003.
- [122] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, S. Krishnan, “NBTI impact on transistor and circuit: models, mechanisms and scaling effects [MOSFETs],” *IEEE International Electron Devices Meeting, IEDM ’03 Technical Digest*, pp.14.5.1-14.5.4, 2003.
- [123] M. Agostinelli, S. Lau, S. Pae, and P. Marzolf, “PMOS NBTI-induced circuit mismatch in advanced technologies,” *IEEE International Reliability Physics Symposium*, pp.171-175, 2004.
- [124] C. Schlunder, R. Brederlow, B. Ankele, W. Gustin, K. Goser, and R. Thewes, “Effects of inhomogeneous negative bias temperature stress on p-channel MOSFETs of analog and RF circuits,” *Microelectronics Reliability*, vol.45, pp.39-46, 2005.
- [125] C. Schlunder, R. Brederlow, B. Ankele, A. Lill, K. Goser, and R. Thewes, “On the degradation of p-MOSFETs in analog and RF circuits under inhomogeneous negative bias temperature stress,” *IEEE Proceedings of International Reliability Physics Symposium*, vol.45, pp.5-10, 2003.

- [126] R. Rodriguez, J. H. Stathis, and B. P. Linder, "A Model for Gate-Oxide Breakdown in CMOS Inverters," *IEEE Electron Device Letters*, vol.24, no.2, pp.114-116, Feb. 2003.
- [127] D. Lee, D. Blaauw, and D. Sylvester, "Gate Leakage Current Analysis and Reduction for VLSI Circuits", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol.12, no.2, pp.155-166, Feb. 2004.
- [128] C.-H. Choi, K.-H. Oh, J.-S. Goo, Z. Yu, "Direct tunneling current model for circuit simulation," *IEEE International Electron Devices Meeting*, Technical Digest, pp.735-738, 1999.
- [129] J. R. Lloyd, "Electromigration for Designers", *Cadence White Paper*, <http://www.cadence.com/whitepapers>, 1999.
- [130] Daniel Young, Aristos Christou, "Failure Mechanism Models for Electromigration," *IEEE Transactions on Reliability*, vol.43, no.2, Jun. 1994.
- [131] J. R. Lloyd, "Reliability Modeling For Electromigration Failure," *Quality and Reliability Engineering International*, vol.10, pp.303-308, 1999.
- [132] L. M. Ting, J. S. May, W. R. Hunter, and J. W. Mcpherson, "AC Electromigration Characterization and Modeling of Multilayered Interconnects," *IEEE IRPS*, pp.311-316, 1993.
- [133] Bharat Sampath, "Electromigration Dependent Mean Time to Failure Calculations," *Univ. of Texas at Arlington*.

- [134] Aurobindo Dasgupta, and Ramesh Karri, "Electromigration Reliability Enhancement Via Bus Activity Distribution," *The 33rd Design Automation Conference*, DAC96, Jun. 1996.
- [135] Mahesh N. Jagadeesan, "Electromigration Analysis for MTTF Calculations," *Univ. of Texas at Arlington*, Oct. 2002.
- [136] Jan M. Rabaey, Anantha Chandrakasan, and Borivoje Nikolic, "Digital Integrated Circuits: A Design Perspective," pp.138-144, Second Edition, *Pearson Education Inc.*, 2003.
- [137] Kaushik Roy, and Sharat Prasad, "Logic Synthesis for Reliability: An Early Start to Controlling Electromigration & Hot-Carrier Effects," *IEEE Transactions on Reliability*. vol.44, no.2, pp.251-255, Jun. 1995.
- [138] JESD63, "Standard Method for Calculating the Electromigration Model Parameters for Current Density and Temperature," *JEDEC STANDARD*, Feb. 1998.
- [139] Richard Blish, Ted Dellin, Steve Huber, etc., "Critical Reliability Challenges for The International Technology Roadmap for Semiconductor (ITRS)," Technology Transfer #03024377A-TR, *International SEMATECH*, Mar. 2003.
- [140] C. K. Hu, R. Rosenbery, H. S. Rathore, D. B. Nguyen, and B. Agarwala, "Scaling Effect on Electromigration in On-Chip Cu Wiring," *IEEE IITC*, pp.267-269, 1999.

- [141] Y. Taur, T. H. Ning, “Fundamentals of Modern VLSI Devices”, *Cambridge University Press*, 1998.
- [142] J. D. Walter, J. B. Bernstein, “Semiconductor Device Lifetime Enhancement by Performance Reduction”, *AVSI Technical Report*, University of Maryland, 2003.
- [143] S. M. Kang, Y. Leblebici, “CMOS Digital Integrated Circuits: Analysis and Design”, *McGraw-Hill College*, 3rd Edition, Nov. 2002.
- [144] J. M. Rabaey, A. Chandrakasan, and B. Nikolic, “Digital Integrated Circuits: A Design Perspective”, Second Edition, *Pearson Education, Inc.*, ISBN 81-78080991-2, pp.214-216, 2003.
- [145] R. S. Muller, T. I. Kamins, M. S. Chan, “Device Electronics for Integrated Circuits”, *John Wiley & Sons*, 3rd Edition, 2002.
- [146] W. Liu, “MOSFET Models for SPICE Simulation, Including BSIM3v3 and BSIM4”, *John Wiley & Sons*, 2001.
- [147] Y. H. Cheng, C. M. Hu, “MOSFET Modeling & User’s Guide”, *Kluwer Academic Publishers*, 1999.
- [148] J. D. Walter, “Methods to Account for Accelerated Semiconductor Wearout in Longlife Aerospace Applications”, *Ph.D. Dissertation*, University of Maryland College Park, 2003.

- [149] M. J. Lakey, "Statistical Analysis of Field Data for Aircraft Warranties", *Proceedings Annual Reliability Maintainability Symposium*, pp.340-344, 1991
- [150] Y. Chen, et al., "A Reliability Evaluation Methodology for Memory Chips for Space Applications When Sample Size is Small", *IEEE International Integrated Reliability Workshop*, Final Report, pp.91-94, Oct. 2003.
- [151] M. Fukuma, H. Furuta, and M. Takada, "Memory LSI Reliability," *Proceedings of the IEEE*, vol.81, issue 5, pp.768-775, May. 1993.
- [152] R. Rodriguez, R. V. Joshi, J. H. Stathis, and C. T. Chuang, "Oxide breakdown model and its impact on SRAM cell functionality," *IEEE International Conference on Simulation of Semiconductor Processes and Devices, SISPAD'03*, pp.283-286, Sept. 2003.
- [153] C. Yao, J. Tzou, R. Cheung, and H. Chan, "Temperature dependence of CMOS device reliability," *IEEE International Reliability Physics Symposium, IRPS'86*, pp.175-182, 1986.
- [154] X. Li, J. D. Walter, and J. B. Bernstein, "Simulating and improving micro-electronic device reliability by scaling voltage and temperature," *IEEE the Sixth International Symposium on Quality of Electronic Design, ISQED'05*, pp.496-502, Mar. 2005.
- [155] W. J. Hsu, B. J. Sheu, S. M. Gowda, and C. G. Hwang, "Advanced Integrated-Circuit Reliability Simulation Including Dynamic Stress Effects," *IEEE Journal of Solid-State Circuits*, vol.27, no.3, pp.247-257, Mar. 1992.

- [156] Boon-Khim Liew, and A. R. Alvarez, "Circuit reliability of hot electron induced degradation in high speed CMOS SRAM," *IEEE Proceedings of the Custom Integrated Circuits Conference*, pp.30.2.1-30.2.4, 1993.
- [157] D. Goguenheim, and A. Bravaix, "Hot-carrier reliability in n-MOSFETs used as pass-transistors," *Microelectron. Reliab.*, vol.38, no.4, pp.539-544, 1998.
- [158] S. Z. Mohamedi, V. H. Chan, J. T. Park, F. Nouri, B. W. Scharf, and J. E. Chung, "Hot-electron-induced input offset voltage degradation in CMOS differential amplifiers," *IEEE 30th International Annual Proceedings of Reliability Physics Symposium*, pp.76-80, 1992.
- [159] J. E. Chung, K. N. Quader, C. G. Sodini, P.-K. Ko, and C. Hu, "The effects of hot-electron degradation on analog MOSFET performance," *IEEE International Electron Devices Meeting*, Technical Digest, pp.553-556, Dec. 1990.
- [160] J. H. Stathis, R. Rodriguez, and B. P. Linder, "Circuit implications of gate oxide breakdown," *Microelectron. Reliab.*, vol.43, pp.1193-1197, 2003.
- [161] R. Rodriguez, J. H. Stathis, B. P. Linder, S. Kowalczyk, C. T. Chuang, R. V. Joshi, G. Northrop, K. Bernstein, A. J. Bhavnagarwala, and S. Lombardo, "The impact of gate-oxide breakdown on SRAM stability," *IEEE Electron Device Letters*, vol.23, issue 9, pp.559-561, Sept. 2002.
- [162] R. Rodriguez, J. H. Stathis, B. P. Linder, S. Kowalczyk, C. T. Chuang, R. V. Joshi, G. Northrop, K. Bernstein, A. J. Bhavnagarwala, and S. Lombardo,

- “Analysis of the effect of the gate oxide breakdown on SRAM stability,” *Microelectron. Reliab.*, vol.42, pp.1445-1448, 2002.
- [163] S. Ikeda, Y. Yoshida, K. Ishibashi, and Y. Mitsui, “Failure analysis of 6T SRAM on low-voltage and high-frequency operation,” *IEEE Transactions on Electron Devices*, vol.50, issue 5, pp.1270-1276, May 2003.
- [164] J. Segura, and A. Rubio, “A detailed analysis of CMOS SRAM’s with gate oxide short defects,” *IEEE Journal of Solid-State Circuits*, vol.32, issue 10, pp.1543-1550, Oct. 1997.
- [165] J. Segura, and A. Rubio, “GOS defects in SRAM: fault modeling and testing possibilities,” *Records of the IEEE International Workshop on Memory Technology, Design and Testing*, pp.66-71, Aug. 1994.
- [166] V. Reddy, A. T. Krishnan, A. Marshall, J. Rodriguez, S. Natarajan, T. Rost and S. Krishnan, “Impact of negative bias temperature instability on digital circuit reliability,” *Microelectronics Reliability*, vol.45, pp.31-38, 2005.
- [167] M. Agostinelli, S. Lau, S. Pae, and P. Marzolf, “PMOS NBTI-induced circuit mismatch in advanced technologies,” *IEEE International Reliability Physics Symposium*, pp.171-175, 2004.
- [168] Y. Leblebici, S.M. Kang, “Hot-Carrier Reliability of MOS VLSI Circuits,” pp. 61, *Kluwer Academic Publishers, Norwell, MA*, 1993. ISBN 0-7923-9352-X.
- [169] B. Kaczer, R. Degraeve, E. Augendre, M. Jurczak, and G. Groeseneken, “Experimental verification of SRAM cell functionality after hard and soft gate

oxide breakdowns,” *The 33rd Conference on European Solid-State Device Research*, ESSDERC’03, pp.75-78, Sept. 2003.

- [170] R. Thewes, R. Brederlow, C. Schlunder, P. Wieczorek, B. Ankele, A. Hesener, L. Holz, S. Kessek, and W. Weber, “MOS Transistor Reliability under Analog Operation”, *Microelectronics Reliability*, vol.40, pp.1545-1554, 2000.