

Aging Assessment and Reliability Aware Computing Platforms

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

donderdag 5 september 2013 om 10:00 uur

door

Yao WANG

Master of Science in Electronic Science and Technology
National University of Defense Technology, China
geboren te Xiangtan, China

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. K. L. M. Bertels

Copromotor:
Dr. S. D. Cotofana

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr. K. L. M. Bertels	Technische Universiteit Delft, promotor
Dr. S. D. Cotofana	Technische Universiteit Delft, copromotor
Prof.dr. L. Fang	National Universiteit of Defense Technische, China
Prof.dr. A. Rubio	Universitat Politecnica de Catalunya, Spain
Prof.dr. J. P. de Gyvez	Technische Universiteit Eindhoven
Prof.dr. P. French	Technische Universiteit Delft
Dr. C. Anghel	Institut superieur d'electronique de Paris, France
Prof.dr. H. J. Sips	Technische Universiteit Delft, reservelid

ISBN 978-94-6186-210-5

Keywords: Reliability, Reliability Aware Computation, Dynamic Reliability Management, Reliability Assessment

Copyright © 2013 Yao WANG

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without permission of the author.

Cover design by Yao Wang.
Printed in The Netherlands

This thesis is dedicated to my parents.

Aging Assessment and Reliability Aware Computing Platforms

Yao WANG

Abstract

Aggressive CMOS technology feature size scaling has been going on for the past decades, while the supply voltage is not proportionally scaled. Due to the increasing power density and electric field in the gate dielectric, the accelerating factors of failure mechanisms in nanoscale Integrated Circuits (ICs) have become more severe than ever. As a result, maintaining IC reliability at the desired level becomes a critical challenge at both design-time and runtime. Addressing the pessimistic reliability landscape outlook over current and future technology nodes, this dissertation investigates reliability-aware design and management techniques to ensure the reliability and quality of IC products. With our special interests on the time-dependent device parameter degradations due to intrinsic failure mechanisms, we focus our discussion on: (i) runtime reliability assessment, (ii) aging degradations, and (iii) mitigation techniques that enable reliability-aware computation. To this end we propose a Dynamic Reliability Management (DRM) framework to combat the aging-induced degradation. In order to achieve a quantitative management, dedicated online aging sensors are employed in the proposed framework to extract dynamic degradation information from circuits. We first propose a unified aging model for the emerging FinFET devices as the physical basis for understanding the underlying aging degradation. Then, we introduce two types of aging sensors, based on threshold voltage and power supply current measurement, respectively, to assist online reliability assessment in DRM systems. Next, we introduce a compensation technique to manage 6T SRAM cell stability under spatial and temporal variations, by threshold voltage modulation using back-gate biasing of independent-gate FinFET devices. We conclude the dissertation by presenting a lifetime reliability modeling and enhancement framework, which demonstrates how to utilize the aging information from dedicated low-level aging sensors to maintain the overall IC health status within prescribed bounds.

Acknowledgments

I acknowledge the help and contributions from many people during my PhD at the Computer Engineering (CE) laboratory of TU Delft. First of all, I would like to give my deepest thankfulness to my supervisor, Dr. Sorin Cotofana, for his sage guidance. He has taught me not only the knowledge and scientific skills, but also the wisdom and philosophy on daily life. He gave me the freedom to explore the research interests on my own, and gave me advice at anytime when I needed it. I can never thank him enough for the countless efforts on correcting and improving my technical writing. I can never forget the many late hours and weekends he sacrificed to work together with us to catch the deadlines. I truly enjoyed and greatly benefited from the past several years of his supervision.

I would like to express my sincere gratitude to Prof. Liang FANG at National University of Defense Technology in China, for his supportive encouragement on my PhD study and the supervision on my scientific initiation. Prof. Fang was the supervisor for my Bachelor's and Master's degree. And he gave me all the convenience to pursue a PhD degree abroad, even though his own research was in urgent need of hands at that time.

I would like to thank Prof. Koen Bertels for serving as my promotor, and also the daily discussions and chatting. He also organized a lot of social events in the CE group to make our social life more colourful. My grateful go to the thesis committee professors as well, for their invaluable feedback and comments despite the tight time schedule. I would like also to thank other faculty members in CE group, Georgi Gaydadjiev, Said Hamdioui, Georgi Kuzmanov, Stephan Wong, Arjan van Genderen, and Zaid Al-Ars, for the interesting talks we had from time to time.

Special thanks are due to the CE secretary Lidwina Tromp for her administrative assistance and generous help. My thanks are also due to Bert, Erik, and Eef, the past and current CE system administrators, especially for their technical support in operating the HPC clusters we used to run our simulations. I would like also to thank the associate coordinator Franca Post from the TU

Delft CICAT office. She helped me to settle down in Delft and took care of the legal documents regarding my stay in The Netherlands.

I would like to thank all the colleagues in the CE group, without whom the life in Delft would not be complete. My special thanks go to past and current officemates formerly in EWI HB. 15.130 and currently in HB. 10.070: Laiq Hasan, Nicoleta Cucu Laurenciu, and Changlin Chen, for creating such a harmonious work place and sharing the numerous technical and non-technical chats with me. Many thanks to Marius Enachescu, George Razvan Voicu, Mihai Lefter, Saleh Safiruddin, Seyab Khan, Mottaqiallah Taouil, Mafalda Cortez, Pavel Zaykov, and Nor Zaidi Haron for their kind help and the interesting talks and jokes we had all the time. I would like to thank Andrew Nelson, Andrew Nelson, Fakhar Anjam, Imran Ashraf, Catalin Ciobanu, Roel Meeuws, and Cuong Pham for the time we spent together.

I truly appreciate the friendship with many great Chinese friends during my stay in the Netherlands. Special thanks to the visiting professors from China, Prof. Donglei ZOU, Dr. Baolan HU, Dr. Fang FANG, Dr. Qijian LIU and Dr. Zhijun DAI. They helped me settle down at Delft and we spent a lot of happy time together. I would like to thank my good friends Jie HU, Yuwei MA, Jinhuan HE, Zongyu LIU, Nannan YU and Wenhua HU, without whom the life would be boring. I will always remember the good times we spent together and the delicious meals we enjoyed. I would like to thank my "landlord", Dr. Chunyang GOU and his wife Ling ZHANG. We were a sort of "family" for the past three years and we got along very well all the time. I would like to thank many wonderful teachers and friends at National University of Defense Technology who created the premises for me to study in the Netherlands.

Finally, I would like to thank my family. I am forever indebted to my parents, for their endless love and unwavering support throughout my life. They sacrificed everything to support me and my brother finish our studies. I also thank my brother for looking after our parents when I was thousands of miles away from home.

Last but not least, thanks to Chinese Scholarship Council, for their financial sponsorship allowing me finish my PhD in The Netherlands.

Yao WANG

Delft, The Netherlands, September 2013

Table of Contents

Abstract	i
Acknowledgments	iii
List of Tables	ix
List of Figures	xi
List of Acronyms and Symbols	xvii
1 Introduction	1
1.1 CMOS Technology Scaling and Its Reliability Trends	1
1.2 CMOS Device Degradation and Failure	4
1.2.1 The Bathtub Curve and Failure	4
1.2.2 Failure Combat in Nanoscale CMOS Designs	5
1.3 Systematic Reliability Management	7
1.4 Dissertation Contributions	9
1.5 Dissertation Organization	10
2 A Framework for Reliability-Aware Design and Computation	13
2.1 A Framework for Reliability-Aware Computation	13
2.2 CMOS Major Aging Failure Mechanisms	15
2.2.1 Bias Temperature Instability	16
2.2.2 Hot Carrier Injection	18
2.2.3 Time Dependent Dielectric Breakdown	19
2.3 Reliability Characterization and Assessment	21
2.4 Reliability Aware Design and Computation	22
2.4.1 Aging-Resistant Architectures	22
2.4.2 Aging-Aware Synthesis	23
2.4.3 Self-Adaptive Tuning	23

2.4.4	Dynamic Task Scheduling & Resource Allocation . . .	24
2.5	Conclusion	24
3	Unified Aging Model for Dynamic Reliability Management	27
3.1	Introduction	27
3.2	FinFET Device	29
3.3	Generalized RD Model and Its 1-D Solutions	29
3.3.1	Generalized RD Modeling Framework	30
3.3.2	Solutions of 1-D Reaction-Diffusion Equation	32
3.4	FinFET Reduced Dimension Effect on NBTI	33
3.4.1	Diffusion Source Limited Size Effect	33
3.4.2	Finite-Oxide Thickness Effect and Oxide-Gate Inter- face Role	36
3.5	HCI Modeling Using RD Model	38
3.6	Model Utilization in Lifetime Reliability Management	42
3.6.1	Lifetime and Aging Definition	42
3.6.2	Degradation Under Random Stress	42
3.7	Results and Discussion	44
3.8	Conclusion	46
4	Aging Sensor Designs for Dynamic Reliability Management	49
4.1	Introduction	50
4.2	V_{th} Based Aging Sensors and DRM System	52
4.3	Circuit Design of V_{th} Sensors	55
4.4	V_{th} Sensors Evaluation	57
4.5	I_{DD} Based Aging Sensors and DRM System	61
4.5.1	I_{DD} Degradation Model Due to Aging	63
4.6	Circuit Design of I_{DD} Sensor	67
4.6.1	The Current Peak Detector (CPD)	67
4.6.2	The Current-to-Time Converter (C2T)	69
4.7	Circuit Performance of I_{DD} Sensors	71
4.8	Conclusion	75
5	Self-Adaptive Compensation Techniques for Independent Gate Fin- FET SRAM	77
5.1	Introduction	78
5.2	Temporal and Spatial V_{th} Variations	79
5.3	ΔV_{th} Impact on SRAM Cell Stability	82
5.3.1	SNM vs. ΔV_{th}	83

5.3.2	SNM vs. ΔV_{th} Modulation	84
5.4	IG-FinFET SRAM Stability Mitigation	85
5.4.1	IG-FinFET Based V_{th} Compensation Scheme	86
5.4.2	V_{th} Compensation Using Supply Leakage Current Monitoring	87
5.5	Simulation Results	91
5.6	Conclusion	100
6	Dynamic Reliability Management - Reliability Assessment	103
6.1	Introduction	104
6.2	Conventional DRM Scheme with V_{th} -Based Sensor	105
6.3	Delay Shift Due to Aging	106
6.4	Time-Sharing Sensing Scheme for Aging Assessment	110
6.5	Delay Shift Calibration under Process Variations	112
6.6	Reliability Metric for DRM System	114
6.6.1	MTTF-Based Lifetime Definition	114
6.6.2	nTTF-Based Lifetime Definition	115
6.6.3	Virtual-Age Definition for Multiple Failure Mechanisms	117
6.7	Experimental Results	119
6.8	Conclusion	122
7	Conclusions and Future Work	123
7.1	Summary	123
7.2	Future Research Directions	126
	Bibliography	129
	List of Publications	137
	Samenvatting	139
	Curriculum Vitae	141

List of Tables

6.1	Delay Shift Estimation for ISCAS85, 89 Circuits.	120
6.2	Delay Shift TSS Based Estimation for ISCAS85, 89 Circuits .	121

List of Figures

1.1	The number of transistor counts per CPU vs. dates of introduction. The solid straight line represents the Moore's Law prediction of the exponential growth of the transistor counts doubling every two years. [Source: Wikipedia]	2
1.2	The power density of Intel microprocessors. The "power density wall" is the fundamental limitation factor that prevents the CMOS technology scale down further.	3
1.3	The "Bathtub Curve" illustrating the failure rate evolution vs. time.	5
1.4	Proposed systematic reliability management framework to combat aging induced degradations.	8
2.1	The Framework of Reliability-Aware Design and Computing Platforms. The major failure mechanisms includes: Negative Bias Temperature Instability (NBTI), Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), Electric Migration (EM), and Thermal Cycling (TC), etc. . . .	14
2.2	Generation of PMOS interface traps under NBTI stress (N_{it} = interface trap).	16
3.1	SOI FinFET Schematic View.	28
3.2	Quasi 2-D cross section for hydrogen diffusion: (a) Triple-Gate (TG) FinFET; (b) Double-Gate (DG) FinFET	34
3.3	Hydrogen concentration profile in finite-size oxide: t_{ox} is the oxide thickness, and gate thickness is considered to be infinite; $\lambda_1 = \sqrt{D_{ox}t}$ and $\lambda_2 = \sqrt{D_{si}t}$	37

3.4	Interface traps generation and hydrogen diffusion in DG-FinFET Channel (top-view): (a) NBTI in DG-FinFET; (b) HCI in DG-FinFET.	39
3.5	Detailed view of interface traps generation and hydrogen diffusion in Planar Structure MOSFET: P is the location of worst case HCI stress, F_L and F_R are the left and right diffusion front on x axis, respectively.	40
3.6	NBTI Under Random Stress	43
3.7	V_{th} Degradation Due to NBTI.	45
3.8	V_{th} Degradation Due to HCI.	46
3.9	V_{th} Degradation Due to the NBTI and HCI Combined Effect.	47
4.1	Schematic of the Proposed DRM System (the Upper Part) and Illustration of the V_{th} Based Aging Sensors (the Lower Part).	53
4.2	Signal Waveform for Degenerated Delay Measurement of Aging Sensors.	54
4.3	(a) NBTI Sensor Circuit Schematic; (b) Measuring Mode Equivalent Circuit Schematic.	55
4.4	(a) HCI Sensor Circuit Schematic; (b) Measuring Mode Equivalent Circuit Schematic.	56
4.5	Temperature and V_{DD} Variation Dependence: the left axis is the V_{th} value from the sensors, and the right axis is the absolute deviation relative to the normal conditions ($T = 27^\circ C$ and $V_{DD} = 1.2 V$) and the data with "cmp" label are for the sensors introduced in [51].	58
4.6	Histogram Plot of the Output V_{th} of NBTI Sensor using Monte-Carlo Simulations.	59
4.7	V_{th} Deviation Relative to Normal Condition ($T = 27^\circ C$ and $V_{DD} = 1.2 V$) with Temperature and Voltage Variations.	60
4.8	Different Measurement Schemes for Degradation Detection: (a) V_{th} sensor scheme; (b) a direct measurement scheme with the proposed I_{pp} sensor (in the right box). The aging indicator I_{pp} of the proposed sensor is taken from the Circuit-Under-Observation (CUO) directly. V_{th} sensor takes the aging indicator V_{th} from the Device-Under-Test (DUT) of the sensor.	62

4.9	Inverter peak current: (a) the circuit model; (b) Voltage-Transfer-Curve (VTC) and the operating regions of PMOS, NMOS transistors; (c) intersections of the output characteristic curves of PMOS, NMOS transistors; (d) the peak power supply current and transistor operating regions.	64
4.10	Peak current of CMOS logic: (a) a general illustration of CMOS network; (b) Equivalent invert circuit for the pull-up network.	65
4.11	Circuit Schematic of the Current-Mode Peak Detector.	68
4.12	Circuit Schematic of the Current-to-Time Converter.	70
4.13	The Transient Waveform of the Voltage across Capacitor C_L	70
4.14	Peak Current Detection at 1GHZ.	71
4.15	Linearity of Peak I_{DD} to Time Converting (left axis) and Error Analysis of Peak Detection (right axis).	72
4.16	The percentage degradations of V_{th} and I_D for all devices in the c499 and c880 circuits - (a) and (b); and the correlations between the percentage degradations of V_{th} and I_D - (c) and (d).	73
4.17	The Time Evolution of the V_{th} and I_D Degradation for 10-year Simulation.	74
5.1	NBTI-Induced V_{th} Degradation Mean Value and Standard Deviation for PTM 32 nm and 20 nm/ FinFET Library [5].	81
5.2	6T SRAM Cell Schematic and Butterfly Curve (PTM 32 nm Technology, $V_{DD} = 0.9 V$).	82
5.3	SNM_L vs. ΔV_{th} Variations for the 6T SRAM Cell Transistors (PTM 32 nm Planar Devices, $CR=\beta_{NL}/\beta_{AXL}=2$).	84
5.4	SNM vs. ΔV_{th} Variations (in the NL, NR Transistors). The contour lines on the bottom plane represent the overall $SNM = \min(SNM_L, SNM_R)$ of the cell.	85
5.5	IG-FinFET 6T SRAM with V_{th} Compensation/Adjustment for PMOS and Pass Gates: $VBPG$ compensates the PMOS NBTI-induced V_{th} degradation, and $tFlex-PG$ bias adjusts the V_{th} of the pass gates to improve the SRAM cell stability.	86
5.6	$Flex-PG$ vs. SRAM Read/Write Stability.	87

5.7	NBTI-induced SRAM cell SNM degradation presented in the $\Delta V_{tPL} \times \Delta V_{tPR}$ plane. The dashed lines with number labels are contour lines for overall SNM of the two nodes. The color lines with cycles are the degraded SNMs after 1 to 9 year(s), and the solid straight red lines represent the SNM time evolution for a given signal probability α at the left node of the SRAM cell.	88
5.8	NBTI Mitigation Using I_{DDQ} Monitor.	89
5.9	A Practical Compensation strategy for NBTI Induced SNM Degradation.	90
5.10	6T-SRAM cell standby leakage and SNM degradation (10-year operation time at $50^\circ C$) using 20nm FinFET Technology with $-10mV$, $0mV$, and $10mV$ global variations, respectively.	92
5.11	6T-SRAM cell leakage distribution of fresh device and aged device(10 years operation at $50^\circ C$) using 20 nm FinFET technology with $-10mV$, $0mV$, and $10mV$ global variations, respectively.	93
5.12	SNM improvement with the VBPG and Flex-PG compensating technique with symmetric double-gate 20nm FinFET technology: $t_{ox1}=t_{ox2}=1.4nm$, $w_{si}=t_{fin}=15nm$	94
5.13	The Applied VBPG Bias with Different Targeted SNM Degradation Margins.	95
5.14	The Cell Leakage Power Consumption versus Different SNM Degradation Targets.	96
5.15	SNM fluctuations under normally distributed NBTI duty cycles α cases. For both cases the deviation $\sigma(\alpha)$ are set to be $0.2\mu(\alpha)$	97
5.16	The required VBPG bias for non-uniform α ratios in SRAM arrays. For both cases the deviation $\sigma(\alpha)$ are set to be $0.2\mu(\alpha)$	99
6.1	Typical DRM scheme with critical path under monitoring. Multiple sensors are required to monitor a single critical path.	106
6.2	CMOS Logic Gates with Pull-up Path(s): (a) 2-NAND; (b) 2-NOR.	108
6.3	Using 2 inverters to substitute the 2-input NAND gate.	108

6.4	Inverter chain with: (a) high to low transition; (b) low to high transition. Only the degradation of inverters with shadow contributes to the NBTI stress induced delay shift.	109
6.5	Time-Sharing Sensing Scheme for Critical Path Monitor.	111
6.6	Relationship Between MTTF and the Weibull Distribution's Parameter.	115
6.7	Illustration of lifetime definition:(a) Logic paths in a pipeline may be not designed with equal delays. A delay guard-band is typically added to combat for aging-induced performance degradation.(b) Lifetime is determined by the path which first eats up all the guard-band. Some paths like St.#1 can have more severe degradation but still meet reliability specification because they have more headroom of delay at time 0.	116

List of Acronyms and Symbols

<i>ABB</i>	Adaptive Body Bias
<i>ASV</i>	Adaptive Supply Voltage
<i>CDF</i>	Cumulative Density Function
<i>CMOS</i>	Complementary Metal-Oxide-Semiconductor
<i>CMP</i>	Chip Multiprocessor
<i>DRM</i>	Dynamic Reliability Management
<i>DFS</i>	Dynamic Frequency Scaling
<i>DVS</i>	Dynamic Voltage Scaling
<i>DVFS</i>	Dynamic Voltage Frequency Scaling
<i>EDA</i>	Electronic Design Automation
<i>EM</i>	Electrical Migration
<i>EOS</i>	Electrical Over Stress
<i>ESD</i>	Electro-Static Discharge
<i>HCI</i>	Hot Carrier Injection
<i>IC</i>	Integrated Circuit
<i>ITRS</i>	International Technology Roadmap for Semiconductor
<i>MOSFET</i>	Metal-Oxide-Semiconductor Field Effect Transistor
<i>MPSoC</i>	Multiprocessor System-on-Chip
<i>MTTF</i>	Mean-Time-To-Failure
<i>NBTI</i>	Negative Bias Temperature Instability
<i>NMOS</i>	N-type Metal-Oxide-Semiconductor
<i>PBTI</i>	Positive Bias Temperature Instability
<i>PDF</i>	Probability Density Function
<i>PMOS</i>	P-type Metal-Oxide-Semiconductor
<i>PVT</i>	Process-Voltage-Temperature
<i>RD</i>	Reaction-Diffusion
<i>RDF</i>	Random Dopant Fluctuation
<i>RTL</i>	Register-Transfer Level
<i>SNM</i>	Static Noise Margin
<i>SRAM</i>	Static Random Access Memory
<i>TC</i>	Thermal Cycling
<i>TDDDB</i>	Time Dependent Dielectric Breakdown
<i>TSS</i>	Time-Sharing-Sensing
<i>TTF</i>	Time-To-Failure

1

Introduction

Since 1960s CMOS (Complementary Metal-Oxide-Semiconductor) device technology has been driven into an aggressive scaling progress. The consistent performance improvement and power consumption reduction due to technology scaling have made CMOS device the dominant technology for Integrated Circuits (ICs). However, as technology scaling enters the deep sub-micrometer regime, CMOS devices are facing a number of quality and reliability issues that have becoming a rising concern to manufacturers and designers. In particular, due to the increasing power density and electric field in the gate dielectric, the accelerating factors of failure mechanisms in nanoscale ICs have become more severe than ever. In addition, due to smaller device dimensions and lower operating voltage, nanoscale ICs have become highly sensitive to environmental fluctuations. As a result, maintaining the reliability of ICs at the desired level becomes a critical challenge to be addressed at both design-time and runtime.

1.1 CMOS Technology Scaling and Its Reliability Trends

Gordon Moore predicted in 1965 that the density of transistors on chip would grow exponentially [77], which is since then known as Moore's Law (See Figure 1.1). Over the past almost five decades, industry has been making full effort to shrink the feature size of Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) devices to reduce the cost per device. The continuous progress of device downsizing has been steadily increasing the performance, decreasing the power consumption, and reducing the manufacturing cost per transistor of IC components. The capability of technology scaling that makes products faster, smaller, and cheaper has led CMOS to become the dominant

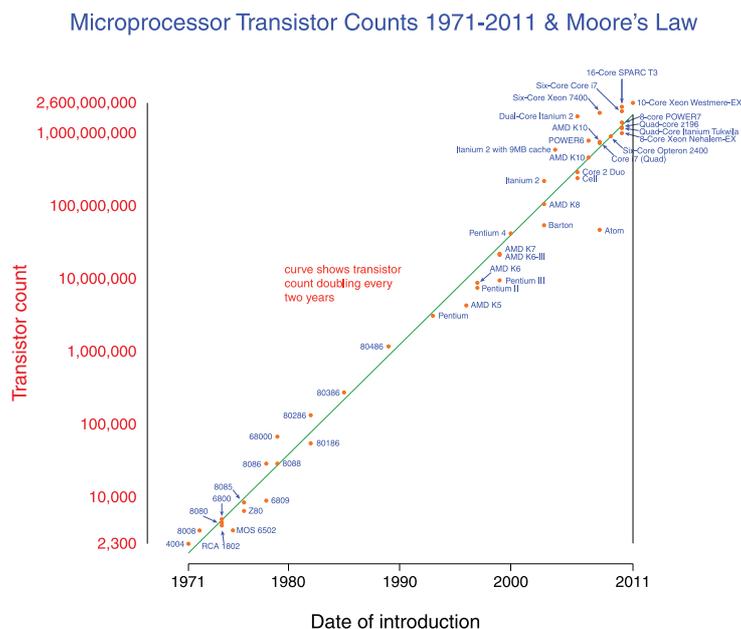


Figure 1.1: The number of transistor counts per CPU vs. dates of introduction. The solid straight line represents the Moore's Law prediction of the exponential growth of the transistor counts doubling every two years. [Source: Wikipedia]

IC technology.

Historically, the fundamental technology scaling guideline is governed by the ideal-scaling principle [29], which specifies that the device dimensions (transistor length L , minimum width W , and oxide thickness t_{ox}) and supply voltage (V_{dd}) should scale down by the same factor α , in order to keep the electric fields in the transistor constant. Hence, this principle is also known as "constant-field-scaling". This ideal-scaling principle has led the CMOS technology to enter the sub-100 nm regime. However, since then, i.e., in the deep submicron regime, the supply voltage (V_{dd}) cannot scale down with the same factor as the device dimensions due to the difficulty of further lowering the threshold voltage (V_{th}). Moreover, as the physical thickness of SiO_2 gate dielectric is scaled beyond 1.2 nm at around 65 nm nodes, the gate leakage current due to quantum mechanical tunnelling becomes significant ($\sim 100 \text{ A/cm}^2$ at 1.0V [70]). Thus, further reduction of the gate dielectric thickness leads to large leakage power consumption. Therefore starting from the 45 nm node, the effective oxide thickness scaling is introduced by using the high- κ gate dielectric to suppress the gate tunnelling current.

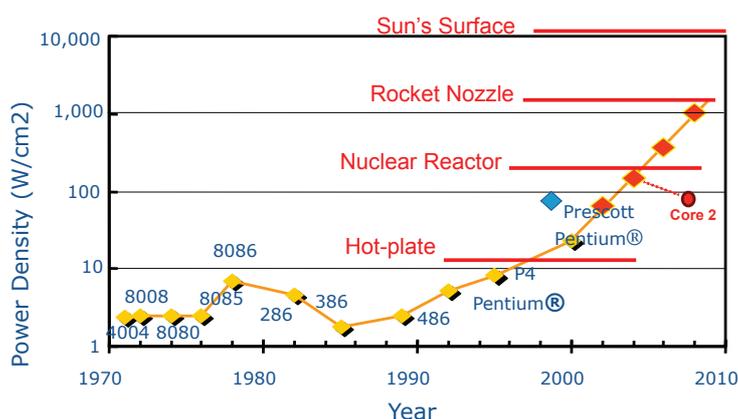


Figure 1.2: The power density of Intel microprocessors. The "power density wall" is the fundamental limitation factor that prevents the CMOS technology scale down further.

Though according to the International Technology Roadmap for Semiconductor prediction [3], the physical gate length of CMOS devices can be expected to scale down to 7 nm before it approaches the physical limitation, the reliability concerns raise dramatically high and become one of the major barrier preventing the technology to scale down further. One of the consequences of non-ideal scaling in the below-100 nm technology nodes is that the electric field in the gate dielectric becomes significantly large, which accelerates the intrinsic failure mechanisms like Time Dependent Dielectric Breakdown (TDDB, [22, 71, 100, 117, 118]), Hot Carrier Injection (HCI, [40, 75, 104, 105, 108]), and so on. As the device dimensions shrink, the number of transistors per area increases, therefore in turn, the on-chip power density increases. The increasing leakage current makes the power dissipation situation even worse. In fact, the temperature in the circuits due to the high power density is now nearly as high as the temperature in the nuclear reactor, as indicated in Figure 1.2. The high temperature accelerates temperature-elevated intrinsic failure mechanisms, e.g., Negative Bias Temperature Instability (NBTI, [28, 72]), Positive Bias Temperature Instability (PBTI, [100]). In a word, the decreasing reliability trend in advanced technology nodes is worsened due to the high gate dielectric electric field and power density in devices and circuits.

As the dimensions become smaller and smaller, the devices become more vulnerable to the process variations, which is considered as another major reliability challenge in advanced technology nodes [20, 34, 72]. In the sub-45 nm era,

the process-variation induced uncertainties of device-level parameters become significantly large relative to the nominal values. These parametric uncertainties create a non-uniform reliability profile in the fresh devices, and affect the time-dependent degradation of device performance as well. As a result, it becomes hardly possible to handle these reliability uncertainties with the prevailing worst-case design assumptions, without introducing a large penalty in terms of area, delay, and power consumption.

1.2 CMOS Device Degradation and Failure

Failure mechanisms in semiconductor can be classified into three categories, namely: intrinsic failures, extrinsic failures, and electrical stress failures [15, 60]. In this dissertation we refer to these terms as follows:

- *intrinsic failures* are those failures that originate from the silicon or die and the processing done in the manufacturing "front end". Potential defects and sources for failure exist in each utilized material and in every process step that alters composition and device features [80].
- *extrinsic failures* are identified with the interconnection and packaging of chips in the manufacturing "back end". Potential failures arise in the steps associated with ensuring that IC contact pads are electrically connected to external power sources and signal processing circuits [80].
- *electrical stress failures* are user-related and event-dependent failures that are mainly caused by Electrical-Over-Stress (EOS) and Electro-Static Discharge (ESD), due to improper handling [80, 87].

1.2.1 The Bathtub Curve and Failure

In an IC product life cycle, the failure rate varies with time because different kinds of failures are dominant in different periods. Empirically, the IC failure rate can be modeled by a "bathtub" curve, which is widely used in reliability engineering across a wide variety of mechanical and electronic components and systems. Figure 1.3 presents an illustration of the failure rates for different technology generations as a function of time.

On the basis of failure rate of products, the bathtub curve can be divided into three stages: the "infant mortality" stage with decreasing failure rate, the "normal operation" stage with near-constant failure rate, and the "wearout" stage

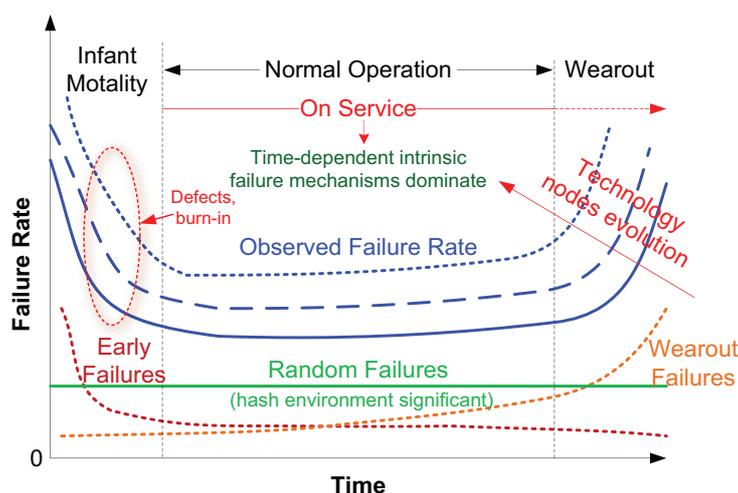


Figure 1.3: The "Bathtub Curve" illustrating the failure rate evolution vs. time.

with increasing failure rate. During ICs' early life the failure rate is high, due to manufacturing defects. In practice, the defective products, as well as the weak ones with a high potential for failure, are eliminated in this stage by a screening or burn-in (i.e., stressing at an elevated temperature, and/or electric field, large/small temperature cycling, and so on) testing process prior to shipment. Once the product reaches customers, the failure rate is relatively small and constant after the burn-in process. The normal operation stage is also known as the "intrinsic failure period" because the time-dependent intrinsic (i.e., "aging") failure mechanisms dominate this stage. It is worth to mention that the random failures are not intrinsic to device but rather to external conditions (e.g., radiation), hence they become significant only in harsh environments. In the late servicing life of IC product, the failure rate increases due to wearout as microscopic defects grow over time and finally take their toll on the product.

1.2.2 Failure Combat in Nanoscale CMOS Designs

To improve the reliability of devices and systems, the most effective way is to reduce the failure rate. For the early life failures, the way to combat is to improve the process technology, in order to reduce the defect density in the products. Alternatively, if the defect sources cannot be eliminated, then screening by burn-in test is conducted to eliminate the defective products. For the aging failure mechanisms, the combating techniques rely on: (i) device engineering

that employ degradation-resistant materials and structures, and (ii) robust circuit and architecture designs that can tolerate device degradation under a given specification.

It is a fact that every system degrades over time and eventually collapses due to the occurrence of a catastrophic failure. However, most systems are designated with a reasonable lifetime expectation. For semiconductor products, the lifetime specification ranges from several years (e.g., consumer electronics) to several decades (e.g., electronic implants). In view of this, the purpose of reliability engineering is to ensure that the failure rate of products during their normal operating life is lower than the maximum acceptable failure rate defined at design-time.

Before the nanoelectronics era, the consequence of aging failure mechanisms was not significant, since the IC lifetime in that period was typically larger than 10 years. However, as we discussed in the previous section, the failure rate due to aging failure mechanisms becomes larger and larger as device dimensions shrink. As a consequence, the reliability related research is now generally focusing on combating aging phenomena in devices and systems, in order to ensure that the wearout does not start before the product is reaching its end of life, according to the lifetime specification it is designed and made for.

Combating the progressive performance degradation induced by aging failure mechanisms raises the following research questions, among others:

- ***How do aging failure mechanisms affect the performance of devices?***
Before any possible technique can be applied to ensure certain IC lifetime reliability requirements, we need physical models of failure mechanisms to guide high level design in reliability assessment, prediction, and tradeoffs. Especially when industry steps into the nanoscale device manufacturing, existing degradation models for aging failure mechanisms should be revised to take consideration of the novel materials and device structures introduced in this era.
- ***How do we assess the IC reliability during its operating life?***
Based on the physical-level understanding of device performance degradation, we need further means to probe the circuit reliability status during the operating life. We note that the understanding of the current reliability status, i.e., reliability assessment, is beneficial to dynamic reliability-aware computing platforms, by means of, e.g., reliability-aware task scheduling and resource allocation.
- ***How do we tolerate or mitigate the aging-induced degradation?***

Different than fault-tolerant techniques, which ensure the system reliability typically by replicating critical components of the system to retain the functionality, the progressive aging degradation can be potentially handled in a less costly way. Since the aging progress is highly dependent on the duty cycle or frequency of usage, one direction to mitigate aging degradation is to balance the workload among components, in order to achieve an overall reliability optimization of the system. Another direction to tolerate aging degradation is to insert pre-defined performance margins at design time. The central question is how to reserve a proper margin, which allows for lifetime reliability while without losing too much performance.

In answering the above questions, as well as targeting on the pessimistic outlook of the reliability landscape in current and future technology nodes, this dissertation investigates reliability-aware design and management techniques to ensure the reliability and quality of IC products. The concepts and challenges of reliability-aware computation are introduced in the next section.

1.3 Systematic Reliability Management

The increased impact of aging effects on circuit performance has drawn a great attention from both industry and academia. Extensive research has been carried on in the related areas, from understanding the very fundamental physical mechanisms of aging degradations (e.g., [13, 14, 23, 31, 36, 55, 84, 85]), modeling device or circuit level performance degradation for CAD tools (e.g., [16, 112, 114]), characterizing/measuring dynamic aging behaviour by online sensors (e.g., [44, 53]), to designing aging-resistant components/architectures to combat/tolerate the aging effects (e.g., [49, 103, 110]). Though each aspect above has been extensively studied, not much work has been done in the direction of identifying a solution able to address the aging degradation issues in a systematic manner.

The traditional approach to handle aging reliability concerns in a digital design flow is to introduce safety margins, which, in turn, reduce the maximum operating frequency and hence decreases the performance. Current practice is to reserve a frequency guard-band of up to 10% – 20% to account for the performance loss due to aging [7, 12, 49]. As technology node advances, the power supply voltage decreases and the relative parameter variations caused by process variations increases, which further interacts with the aging degradation and leads to significant performance variations [20, 72]. As a consequence, the

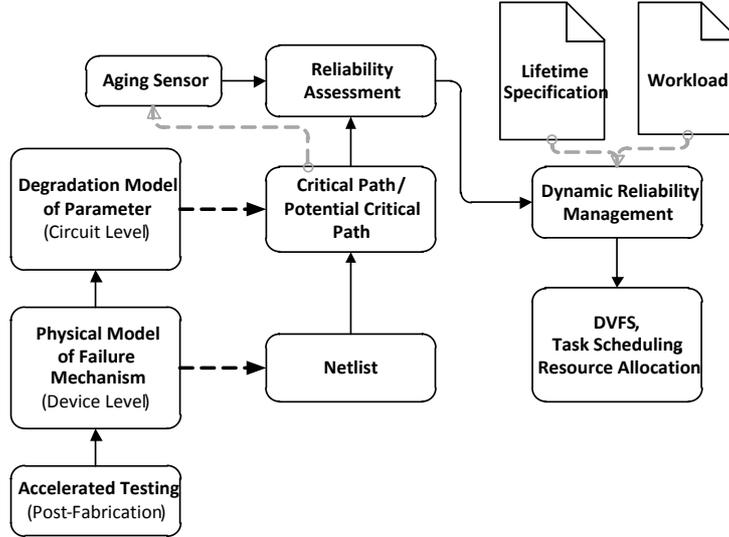


Figure 1.4: Proposed systematic reliability management framework to combat aging induced degradations.

safety margins have to be set even more conservatively, which can minimize or even eliminate the advantage introduced by a smaller technology node [65]. As a result, the tremendous effort and money spent on technology development are partially or fully wasted.

In order to bring the performance back and fully take advantage of the newer technology node, the safety margins reserved for aging degradation must be tightened. To achieve this goal, in this dissertation, we propose a bottom-up dynamic reliability-aware computing platform able to address reliability related issues in a systematic way. The main principle behind our systematic solution is to utilize low-level in-field aging information to guide the high-level aging combating/mitigation techniques. The in-field real-time collected aging information allows us to perform a more efficient reliability management when compared with the conventional solutions [49, 103, 110].

Figure 1.4 depicts our systematic reliability assessment and management framework which targets at the run-time combat of aging induced degradations. Our proposal relies on device level physical model of failure mechanism, which is extracted from the post fabrication accelerated testing. After that, the circuit level degradation model associated with a specific performance parameter, e.g., delay, is built upon the device level physical model. A dedicated aging sensor is utilized to dynamically extract the aging-indicating parameter,

e.g., threshold voltage V_{th} , from the device or circuit. The aging sensor raw data output is then processed by the Reliability Assessment module in order to extrapolate circuit information into an aging status that can be further utilized in failure time prediction and/or reliability aware resource management. Dynamic Reliability Management (DRM) module make use of the obtained information to further guide the high-level aging mitigation techniques (e.g., Dynamic Voltage/Frequency Scaling, DVFS) or to perform a reliability-aware task scheduling and resource allocation.

Most of the existing DRM proposals [19, 97, 119] perform a "blind" optimization for reliability, since they do not rely on dedicated aging sensor no actual aging information can be provided. When compared with those proposals, our proposal utilize the aging sensor to dynamically extrapolate the aging information from the circuits. With the collected aging profiling data from the sensors, a more proactive reliability management policy can be adopted and the safety margin reserved for the aging effects can be speculatively updated at runtime. As a result, the performance loss due to the conservative margins at design-time can be eliminated and the technology potential can be better utilized. In addition, our proposal enables application specific execution scenarios that can trade performance, energy consumption for reliability. Furthermore, our proposal can provide an alarm signal before a failure actually happens in the circuit, which is highly desirable in reliability-critical applications.

1.4 Dissertation Contributions

The focus of this dissertation is on reliability assessment and dynamic management for reliability-aware computing platforms. In this area of research, it makes the following contributions:

- A unified aging model of NBTI and HCI degradation in emerging FinFET devices towards lifetime reliability management for nanoscale MOSFET circuits.
- A dynamic reliability management framework with two types of aging sensor designs, namely threshold voltage (V_{th}) based and power supply current (I_{DD}) based aging sensor, respectively.
- An investigation on the impact of spatial and temporal variation on the stability of SRAM arrays and a mitigation technique applicable for independent-gate devices.

- A lifetime reliability assessment framework for combinational logic with a time-sharing aging information sensing scheme from low-level sensors.
- A "virtual age" based system reliability metric to overcome the shortcomings of conventional mean-time-to-failure (MTTF) for reliability optimization.

1.5 Dissertation Organization

The remainder of this dissertation is organized as follows.

Chapter 2 gives an introduction on the framework of reliability-aware computation platform, from the major aging failure mechanisms in the nanoelectronics era to the state of the art in reliability assessment and reliability aware computing techniques. Specifically, it provides a survey on the physical model of major aging failure mechanisms, existing aging sensor designs, circuit level degradation models, design-time reliability enhancement proposals, and runtime reliability management schemes.

Chapter 3 introduces a unified reliability model of Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) for double-gate and triple-gate FinFETs, towards a practical reliability assessment method for future FinFETs based circuits. The model is based on the reaction-diffusion theory and it is extended to cover the FinFET specific geometrical structures. Apart of introducing the reliability model we also investigate the circuit performance degradation due to NBTI and HCI in order to create the premises for its utilization for assessing and monitoring the Integrated Circuits (ICs) aging process. Simulation results suggest that our model characterize the NBTI and HCI process with accuracy and it is computationally efficient, which makes it suitable for utilization in reliability-aware architectures as reliability prediction/assessment kernel for lifetime reliability management mechanisms.

Chapter 4 proposes two types of aging sensor designs, based on threshold voltage (V_{th}) and on power supply current (I_{DD}) measurement, respectively, to assess the reliability status from device/circuit. The V_{th} -based aging sensor is highly tolerant to process, voltage, and temperature variations, which is highly desirable for accurate reliability assessment. However, the I_{DD} -based aging sensor can extract the amalgamated effect of various aging mechanisms, e.g., NBTI, HCI, from a large circuit block, which can significantly reduce the required number of aging sensors for acquiring degradation information from

a large system. Besides the circuit designs for the aging sensors, two Dynamic Reliability Management (DRM) system schemes associated with these two types of sensors are presented as well. Both of our sensor designs are verified by simulations with Cadence tools using TSMC 65 nm technology library. The simulation results indicate that the V_{th} -based aging sensor has a very low process, supply voltage, and temperature (PVT) sensitivity, which outperforms the accuracy of the NBTI and HCI sensors from prior work under PVT-variation circumstances; and the power supply current I_{DD} exhibits a similar aging rate as the threshold voltage for the entire circuit lifetime, but with a better sensitivity towards the End-of-Life (EOL), which demonstrates the validity and practical relevance of the proposed I_{DD} aging monitoring framework.

Chapter 5 investigates the influence and mitigation of NBTI degradation and random process variations on the stability of the FinFET based 6T-SRAM cell. The contributions of transistor threshold voltage variations (ΔV_{th}) on the stability of the SRAM cell are thoroughly examined by means of SPICE simulations. Different biasing schemes for compensation at different transistors are investigated, and the optimal scheme is selected. A mitigation method for memory stability management under spatial and temporal variations is demonstrated. By taking advantage of the independent-gate FinFET device structure, threshold voltage adjustment is performed by back-gate biasing. The proposed technique allows for a practical compensation strategy able to preserve the SRAM cell stability while balancing performance and leakage power consumption. We evaluate the impact of our proposal on the SRAM cell stability by means of SPICE simulations for 20nm FinFET devices. Simulation results indicate that the proposed technique can effectively maintain stability of an SRAM array within the desired range during its operational life under both spatial and temporal variations, hence improve the system performance and reliability. Our method allows for maintaining the Static Noise Margin (SNM) degradation of SRAM cells under a certain range, e.g., 2% of fresh device after 1 year operation, which is about 55.56% improvement when compared with the 4.5% degradation corresponding to the uncompensated case.

Chapter 6 presents a lifetime reliability modeling and enhancement framework, which demonstrates how to utilize the aging information from dedicated low-level aging sensors to extrapolate the overall system health status. We first propose a path delay shift model to link up the measured aging information with the circuit degradation. Then we propose a Timing-Sharing Sensing (TSS) method for V_{th} -based DRM to sample the dynamic activity ratio from the circuits under monitoring. Furthermore, we introduce a system-level reliability metric "Virtual Age", instead of the commonly used Mean-Time-to-

Failure (MTTF) metric, for dynamic reliability management. We evaluate our approaches by conducting SPICE simulation on a couple of ISCAS 85 & 89 benchmark circuits and the results have indicated that the proposed path delay shift estimation model and TSS scheme can predict/assess the circuit performance degradation, i.e., circuit path delay shift, with an acceptable accuracy margin no more than 5.03%.

Finally, **Chapter 7** concludes our work and provides some directions for future research.

2

A Framework for Reliability-Aware Design and Computation

Reliability aware computation involves a wide range of aspects. In this chapter, we give an overview of the reliability-aware computation framework. Specifically, we first give a review on the major aging failure mechanisms in nanoelectronics era, and then present the state of the art in reliability assessment and reliability aware computation, which includes details on existing aging sensor designs, circuit level degradation models, design-time reliability enhance proposals, and runtime reliability management schemes.

2.1 A Framework for Reliability-Aware Computation

The reliability-aware computation related work can be depicted by a framework as presented in Figure 2.1. The reliability aware computing is fundamentally built upon the solid understanding of physical and thermodynamic process of aging failure mechanisms. These processes are described by device parameter degradation models. Based on the device-level degradation models, circuit-level degradation models can be further derived to guide high level reliability-aware synthesis, reliability assessment, and reliability-aware task scheduling. In addition, aging-resistant architectures and self-adaptive techniques are widely proposed (e.g., [8, 26, 94]) for reliability-aware computing as well.

Usually, IC lifetime requirements are mostly made based on worst-case assumptions, which leads to highly conservative margins on technology parameters, resulting in the under utilization of the technology potential. To make better use of the technological improvement, this pessimistic assumption should

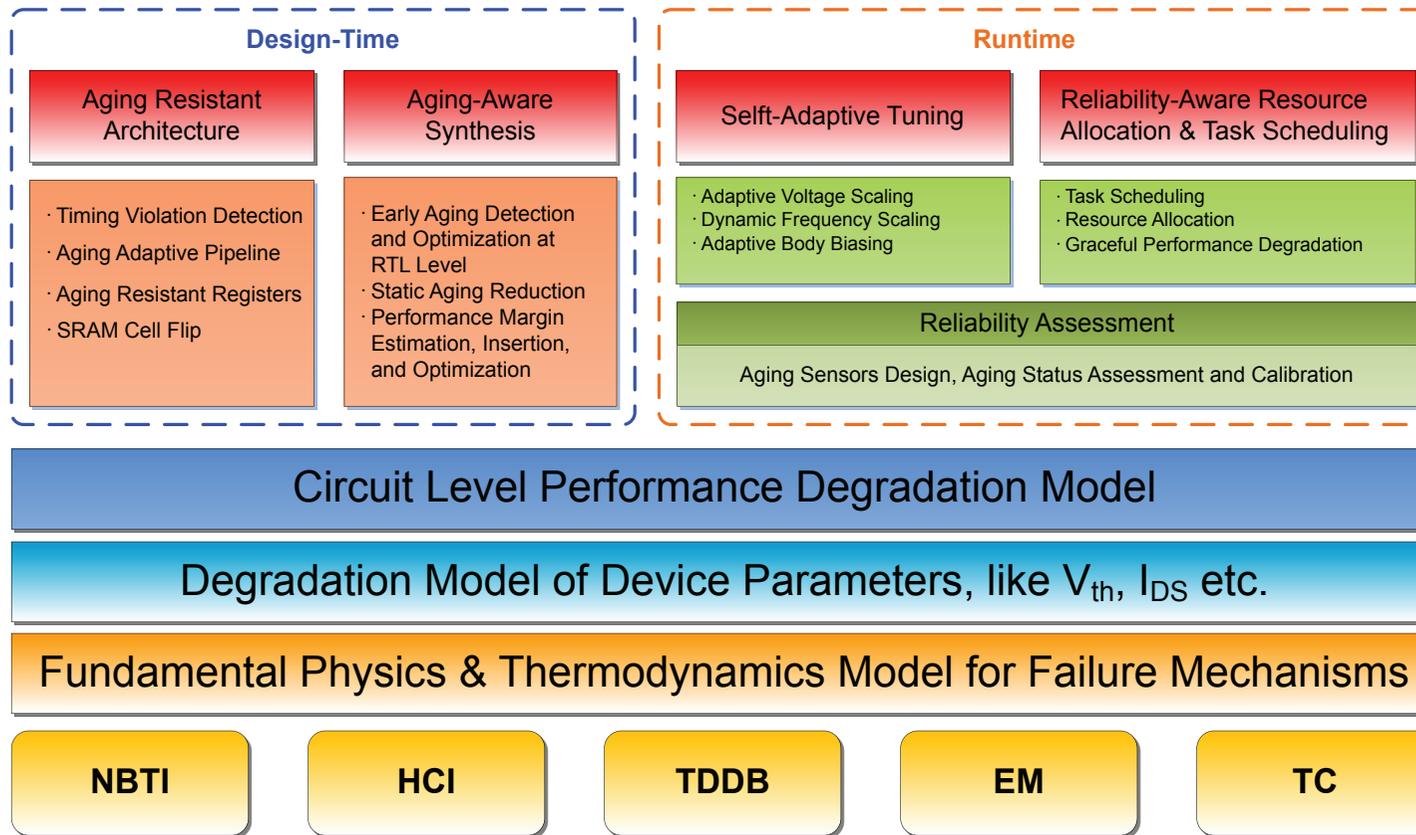


Figure 2.1: The Framework of Reliability-Aware Design and Computing Platforms. The major failure mechanisms includes: Negative Bias Temperature Instability (NBTI), Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), Electric Migration (EM), and Thermal Cycling (TC), etc.

be relaxed and combined with a Dynamic Reliability Management (DRM) framework that relies on aging sensors able to provide a reliability status assessment on the ICs.

Among the high level reliability-aware techniques, aging-resistant circuit and architecture design and reliability-aware synthesis are conducted at design-time. On the contrary, self-adaptive tuning and reliability-aware task scheduling techniques handle aging degradation at runtime. The real challenge of reliability-aware computation is how to satisfy the lifetime specification with minimum overhead of area, delay, and power consumption. To be more specific, we need to understand how much degradation aging can induce in practice under different conditions, how can we evaluate the existing degradation, and how can we allocate the necessary resource to combat the aging induced degradation. Addressing these challenges at runtime, DRM techniques have more flexibility than design-time techniques, since adjustment according to different IC aging conditions can be performed at the individual, even the component scale. Hence, there is a higher possibility for runtime techniques to achieve a better optimization than design-time techniques by utilizing less resources. With our special interests on the time-dependent device parameter degradations due to intrinsic failure mechanisms, we focus our discussion on runtime reliability assessment of aging degradations, and on mitigation techniques that enables reliability-aware computation. However, this does not restrict the generality of our work, such as device- and circuit-level reliability models, which can be also applied to the design-time reliability assessment and optimization.

In the reminder of this chapter, we walk through an overview of the existing work related to the reliability-aware computation in a bottom-up way. In the next section, we give a review on the major aging failure mechanisms in the nanoelectronics era, including the physical mechanism of the aging effects, the damage it causes to the circuit, and the mathematical model we use to describe the time evolution of the failure mechanism.

2.2 CMOS Major Aging Failure Mechanisms

Aggressive technology scaling in the past decades has made CMOS devices very vulnerable to aging degradations. There are a number of physical failure mechanisms that can affect the reliability of CMOS devices. According to the ITRS 2011 guideline on process integration, devices and structures, the major reliability concerns in the near future include [3, 69]:

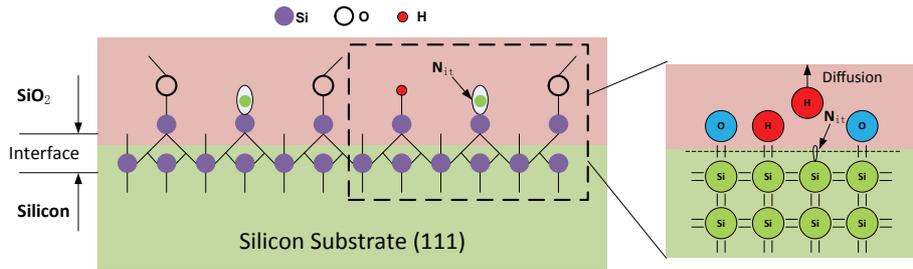


Figure 2.2: Generation of PMOS interface traps under NBTI stress (N_{it} = interface trap).

- Negative/Positive Bias Temperature Instability (NBTI/PBTI), Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), and Random Telegraph Noise (RTN) in scaled and non-planar devices;
- Electro-Migration (EM) and Stress Voiding in scaled interconnects;
- Increasing statistical variations of intrinsic failure mechanisms in scaled and non-planar devices.

In the rest of this section, we give a brief introduction on the underlying physics of the most important aging mechanisms, namely BTI (including NBTI and PBTI), HCI, and TDDB, in order to provide the reader with a basic understanding of the aging effects and how they evolve with technology.

2.2.1 Bias Temperature Instability

Negative Bias Temperature Instability (NBTI) is considered to be one of the most critical reliability threats to device in the nanoelectronics era [28, 72]. NBTI is prominent in PMOS devices along the entire channel when negative gate-to-source voltage is applied, resulting in the generation of interface traps (N_{it}) at the Si-SiO₂ interface. The accumulation of interface traps causes a gradual shift of key transistor parameters, such as threshold voltage (V_{th}), linear (I_{dlin}) and saturation (I_{dsat}) drain current, and transconductance (g_m).

Figure 2.2 illustrates the interface traps generation at the Si-SiO₂ interface (to be more precisely, it is the substrate-oxide interface). In the fabrication process, the dangling bonds (i.e., unsatisfied silicon valence electrons) at the Si-SiO₂ interface are passivated with hydrogen atoms. Under certain circumstances (mainly due to elevated temperature), inversion layer holes from the channel can tunnel into the gate oxide, break the Si-H bond leaving behind an

interface trap. Consequently, the free hydrogen particles are diffused/drifted away from the Si-SiO₂ interface. Since the PMOS device operates with a negative gate bias, the electric field in the gate dielectric layer is directed away from the Si-SiO₂ interface. This electric field drives the holes tunnel from the channel to the oxide, which makes NBTI for PMOS devices a more severe issue than the PBTI associated with NMOS devices. However, PBTI is still an important reliability issue when high- κ materials are utilized for gate dielectric [100].

Despite of the fact that the NBTI effect has been extensively studied in recent years, there is no general agreement on the physical fundamentals of the NBTI kinetics. One of the most popular NBTI models is the Reaction-Diffusion (RD) model, which attributes the phenomenon to a thermally activated reaction of holes with Si-H bonds at the Si-SiO₂ interface. The RD mechanism was first proposed in 1977 by Jeppson and Svensson in [45], and they found out that the NBTI-driven shift of the P-MOSFET threshold voltage depends on the applied gate voltage, temperature, and stress time. Alam et al. extended this model in [13,14,54,55,68] to capture the NBTI-induced parameter shift (mainly, V_{th}) in nanoscale CMOS technologies. According to the RD model, the threshold voltage shift ΔV_{th} can be expressed as:

$$\Delta V_{th} = A \exp\left(-\frac{E_a}{kT}\right) \exp(\gamma V_{gs}) t^n, \quad (2.1)$$

where A is a technology-dependent constant, E_a is the activation energy (typically ~ 0.1 eV), k is the Boltzmann constant, T is the absolute temperature, V_{gs} is the applied gate voltage, γ is a gate voltage dependent coefficient, t is the time in seconds, and n is the NBTI time evolution exponent. Unfortunately, Eq. (2.1) is only valid for static (also referred as "DC") stress conditions. The dynamic voltage stress condition is taken into consideration in [59]. A compact NBTI stress model is proposed in [112], where a more advanced analytical NBTI model with parameters extracted for a 65 nm CMOS technology is presented.

The RD model is widely used in academia, however, it is not able to fully capture all the features of NBTI phenomena. One major criticism is that the RD model predicts a universal recovery of the NBTI induced damage when the applied negative gate bias is removed. This is in a big contrast with experimental measurements which indicate that NBTI recovery cannot be the diffusion-limited process as suggested by the RD model [36]. As an alternative, models using fast hole trapping and detrapping in NBTI-generated and pre-existing traps at the Si-SiO₂ interface or in the oxide are proposed to explain the fast

NBTI recovery effect. In [84, 85] Parthasarathy et al. proposed a model which is combining the interface generation and hole trapping together to explain the recovery characterization of NBTI effect. Later, Grasser and Kaczer et al. [37, 88] presented a trapping model demonstrating how a unified field and temperature acceleration account for both the stress and recovery phase. Experimental measurements on very small devices (< 100 nm) has revealed that the NBTI recovery behaviour takes place in discrete steps [36]. The properties of these discrete steps are not consistent with a diffusion-limited process, but rather with the capture and emission of individual holes.

The hole trapping/detrapping model seems to be closer to the physical fundamentals, however, there is criticism on it as well [68]. First, the hole trapping model cannot predict the power law time exponent as RD model does. Second, the RD model reveals a robust universality of of NBTI degradation under dynamic (also referred as "AC") stress, which has not been predicted before. The contention on the NBTI physics process is still going on, as none of the existing models can capture all the NBTI features. Moreover, there is a lack of compact models formulated in terms of transistor design parameters for design-time and runtime reliability evaluation and assessment.

2.2.2 Hot Carrier Injection

Hot Carrier Injection (HCI) was a major reliability concern in 1980s [40, 104, 105]. Later, it became less dominant when the supply voltages were scaled down and graded drain junctions were introduced. However, HCI draws attention again in nanoscale electronics due to the increasing internal electric fields present in aggressively scaled devices [75, 108].

HCI occurs when an electron or a hole gains sufficient kinetic energy to inject itself from the channel into the gate dielectric, leaving behind an interface trap just as NBTI does. Consequently, the transistor parameters, e.g., V_{th} and so on, shift gradually over time. The term "hot" refers to the fact that the carriers are accelerated to a considerably higher energy than the thermal energy of the surrounding lattice. The hot carriers tunnelling into the gate dielectric can show up as gate current. However, the substrate current, which is the opposite component of the HCI induced gate current, is usually collected to evaluate the HCI stress induced damage, because of the easier measurement.

HCI is one of the most thoroughly investigated aging effects. The very basis of most HCI models are build on the "lucky electron" model. This model was first introduced in the work [40, 104] and it is based on the following

assumptions: (i) the carrier energy is gained solely through the drain lateral electric field acceleration; (ii) the main energy relaxation process is phonon scattering. According to this model, the device parameter degradation can be expressed as [100]:

$$\begin{aligned} \Delta P &\propto L_{eff}^{\alpha} \times \Delta N_{it}(t) \\ &\approx L_{eff}^{\alpha} \left[t \times \frac{I_d}{W} \times \left(\frac{I_{sub}}{I_d} \right)^{m \gamma} \right]^n \exp \left(\frac{E_a}{kT} \right), \end{aligned} \quad (2.2)$$

where ΔP is the change in the device parameter (e.g., ΔV_{th}), L_{eff} is the effective channel length, W is the device width, ΔN_{it} is the generated interface traps due to HCI, t is the time in seconds, I_d is the drain current of fresh device, I_{sub} is the substrate current of fresh device, k is the Boltzmann constant, T is the absolute temperature, E_a is the HCI activation energy (typically around -0.05 eV), α is a technology related constant, m is a constant given by ϕ_{IT}/ϕ_{II} , where ϕ_{IT} is the critical energy for carriers to create an interface trap (≈ 3.7 eV, [40]) and ϕ_{II} is the impact ionization threshold energy (≈ 1.3 eV, [100]), and n is the power-law time exponent for HCI induced degradation time evolution.

2.2.3 Time Dependent Dielectric Breakdown

Time Dependent Dielectric Breakdown (TDDB) is a result of high operating electric fields in the gate dielectric of MOSFET devices. The breakdown is caused by the formation of a conducting path through the gate oxide to substrate due to electron tunnelling current, when MOSFETs are operated close to or beyond their specified operating voltages.

When the gate dielectric is subjected to electrical stress, structural defects are continuously generated in the oxide bulk and at the interface as well. With the defect number increases, soft breakdown (SBD) is formed, which leads the dielectric to a partial loss of the insulativity. The accumulation of SBD may cause a final hard breakdown (HBD) of dielectric, which leads the dielectric to a complete loss of its functionality. The physical consequences of the dielectric degradation include [100]: positive charge trapping; generation of neutral electron traps; generation of Si-SiO₂ interface states; increase of the gate leakage current, etc.

Historically, a number of TDDB models have been widely used to predict the time-to-breakdown (i.e., when the HBD occurs) t_{BD} due to the TDDB failure mechanism. The most important models are the thermochemical model [71], the anode-hole-injection model [22], and the voltage model [117, 118].

The thermochemical model is also known as the E model, holds true for gate oxide thickness is greater than 4 nm [1]. The cause of TDDB under low electric field (< 10 MV/cm) is due to field-enhanced thermal bond breakage at the Si-SiO₂ interface. The time-to-breakdown t_{BD} can be expressed as [71]:

$$t_{BD} = A_0 \exp(-\gamma E_{ox}) \exp\left(\frac{E_a}{kT}\right), \quad (2.3)$$

where γ is the field-acceleration parameter, E_{ox} is the electric field in the oxide, E_a is the activation energy of TDDB, and A_0 is a process/material dependent coefficient.

The Anode-Hole-Injection model is also referred to as the $1/E$ model. According to this model, the TDDB damage is assumed to be due to current through the dielectric by Fowler-Nordheim (F-N) conduction. F-N injected electrons (from the cathode) cause dielectric impact ionization damage as they accelerate through it. The time-to-breakdown t_{BD} is expected to exhibit an exponential dependence on the inverse of electric field [22]:

$$t_{BD} = \tau_0(T) \exp\left[\frac{G(T)}{E_{ox}}\right], \quad (2.4)$$

where $\tau_0(T)$ is a temperature dependent prefactor, and $G(T)$ is a temperature dependent field acceleration parameter for the $1/E$ model. In contrast with the E model, the $1/E$ model has been proved to provide a good fit with experimental data in situation when a high electric field is applied.

Both the E model and $1/E$ model can only fit a limited range of the electric field. Nevertheless, the applicability of these two models does not appear to be valid for gate oxide thickness smaller than 4 nm. It was found that the breakdown of the ultra-thin oxides show an exponential dependence on voltage rather than on the electric field. In such cases, the time-to-breakdown t_{BD} can be expressed as [117, 118]:

$$t_{BD} = A_0 \exp(-\beta V) \exp\left(\frac{E_a}{kT}\right), \quad (2.5)$$

where A_0 is a material and process dependent factor, β is the voltage acceleration parameter, and V is the applied voltage.

BTI, HCI, and TDDB are considered as the most important aging effects in the current and future technology nodes. The review above describes the physical fundamentals and the most popular time-evolution models associated to them. However, as novel devices emerge in the nanoelectronics era, these models

should be revised to accommodate the novel geometry structures and materials introduced in those devices. In Chapter 3 we give further insight on the RD model, and we extend it to describe the time evolution process of NBTI and HCI effects in FinFET devices. In the next section, we give an introduction on the reliability characterization and assessment techniques, which are utilized to understand and evaluate the degradation process due to the aging effects.

2.3 Reliability Characterization and Assessment

Up to date most of the high level proposals for reliability-aware computation perform dynamic reliability management without relying on dedicated aging sensors to extract reliability status from circuits [19,27,42,94,97,101]. Though they can gain a statistical reliability improvement across a large population of products, those proposals might fail on particular individual ICs due to the spatial and temporal variability. To overcome this limitation, in this dissertation, we propose a reliability-aware computing platform with the support of reliability assessment by dedicated online aging sensors.

In the recent past, a number of approaches for aging/reliability characterization or monitoring have been reported. In [53], Kim et al. introduce an on-chip aging monitor for NBTI. High resolution of degradation measurements can be achieved by detecting the beat-frequency from a pair of ring oscillators. Keane et al. further extend this idea to an "all-in-one" sensor for BTI, HCI, and TDDB degradation measurement [44]. However, the large area overhead of their design makes their approach suitable only for degradation characterization, not for online reliability assessment.

Karl et al. propose compact in-situ sensors for monitoring NBTI and TDDB, respectively, in [48, 96]. These sensors work in the sub-threshold region with leakage current to increase the sensitivity. Even though they require a small area overhead, these sensors are sensitive to process, voltage, and temperature variations. Agarwal et al. propose aging sensor designs integrated inside a flip-flop to detect delay violation(s) in [10, 11]. These designs are relatively small and can be potentially included in many chip flip-flops. However, this kind of sensor can only check delay violation in a static or quasi-static time window ("guard-band"), and thus no quantitative aging information can be collected.

The previous analysis clearly indicates that existing aging sensors either have large area overhead, which makes them not suitable for online reliability assessment; or cannot provide quantitative aging information at all. In order to achieve our goal of implementing a proactive DRM system, we propose two

types of aging sensors in Chapter 4 based on threshold voltage (V_{th}) and power supply current (I_{DD}) measurement, respectively.

2.4 Reliability Aware Design and Computation

Addressing the increasing concerns on reliability and aging degradation in nanometer IC era, reliability-aware computation become a hot topic in recent years. Tremendous effort has been given on reliability-related design and computation. The proposals can be divided into four categories, namely: aging resistant architecture, aging-aware synthesis, self-adaptive tuning techniques, and dynamic task scheduling and resource allocation.

2.4.1 Aging-Resistant Architectures

At the beginning, most work on aging resistant architecture focused on design methodologies rather than on specific aging mitigation techniques. A first step towards addressing this issue called "RAMP" is proposed in [97], which is a micro-architecture level model that allows for performance boosting within an acceptable reliability margin. In [103], Sylvester et al. propose an adaptive self-healing architecture named "ElastIC" to deal with the extreme conditions in a multiple-core processor subject to huge process variations, transistor degradations at varying rates, and device failures. ElastIC is based on aggressive runtime self-diagnosis, adaptivity, and self-healing. However, this work is just a conceptual investigation on architecture-level methodologies to combat process variations and aging degradation, and no specific implementations of the combating techniques are presented. Similarly, in [110] Tiwari et al. propose a framework named "Facelift" to hide the performance degradation due to aging through aging-driven application scheduling, and Adaptive Supply Voltage (ASV) or Adaptive Body Bias (ABB). Recently, some architecture-/microarchitecture-level aging-resistant techniques are proposed. In [58] Kumar et al. use a data flipping technique to recover the Static Noise Margin (SNM) of the SRAM cell. However, the performance and area overhead introduced by the data flipping technique is very high. In [8], Abella et al. propose an NBTI-aware processor named "Penelope", which integrates several strategies to mitigate NBTI. The main idea of Penelope is to enhance the recovery effect of NBTI during idle time for combinational logic blocks and memory-like blocks, respectively. This method is economic efficient but interrupts the normal functionality of the processors.

2.4.2 Aging-Aware Synthesis

Aging-aware synthesis mainly refers to degradation mitigation and/or optimization techniques adopted in EDA tools at design-time. In [18] Bild et al. present a technique for the minimization of the NBTI induced performance degradation by internal node control. In this technique, the input signals of individual gates are directly manipulated to prevent the static NBTI fatigue. More specifically, to eliminate static NBTI stress on all the PMOS transistors in a circuit, the outputs of most gates must be forced high. Gates feeding only into the lower PMOS transistors of NOR gates are the exception. In [115] Wang et al. propose two gate replacement algorithms, which together with optimal input vector selection can simultaneously reduce the leakage power and mitigate NBTI-induced degradation. In [56] Kumar et al. present a methodology to estimate NBTI induced delay degradation at the Register Transfer Level (RTL) by signal correlations analysis. In this way, aging-aware optimization can be conducted at RTL code level.

2.4.3 Self-Adaptive Tuning

Usually, IC lifetime requirements are mostly determined based on worst-case assumptions, which leads to highly conservative margins on technology parameters, resulting in the under utilization of the technology potential. To make better use of the technological improvement, Dynamic Reliability Management (DRM) frameworks are proposed to diminish the performance loss due to the design-time pessimistic assumption. The most frequent techniques adopted in DRM frameworks are Dynamic Voltage Frequency Scaling (DVFS) and Adaptive Body Biasing (ABB) [27, 49, 74, 94]. In [27] Das et al. presented a Dynamic Voltage Scaling (DVS) technique called "Razor", which incorporates an in situ error detection and correction mechanism to recover from timing errors. Although their work is not dedicated to aging-aware computation, their proposal can be easily adapted to the related area. In [94] Shah et al. further extended the DVS technique into a Built-In Proactive Tuning (BIPT) which the basic principle is that each circuit block can autonomously tune its performance according to its own degree of aging. In [49] a self-adaptive architecture is proposed to address transistor aging by DVFS as the devices age. In [74] Mintarno et al. present a framework and several control policies for optimizing the dynamic control of various self-tuning parameters over lifetime in the presence of circuit aging.

2.4.4 Dynamic Task Scheduling & Resource Allocation

Besides the self-adaptive tuning techniques, dynamic task-scheduling and resource-allocation techniques are proposed as well to deal with aging degradation in the circuits. Huang et al. propose an algorithm to perform task allocation on Multiprocessor System-on-Chip (MPSoCs) platforms in order to minimize the energy consumption while satisfying a given lifetime reliability constraint in [41, 42]. However, their work is inherited from thermal-aware task scheduling and allocation algorithms, hence it might not result in reliability optimized results due to the underestimation of other aging accelerating factors, e.g., electric field, duty-cycle, and so on. In [101, 102] Sun et al. propose an NBTI aware system workload model and a Dynamic Tile Partition (DTP) algorithm to balance workload among active cores while relaxing the stressed ones on Chip Multiprocessor (CMP) platforms.

Our analysis clearly indicates that reliability aware computation is attracting more and more attention from both academia and industry. Within the various possibilities to handle aging degradation related issues, this dissertation emphasizes the effort on the runtime techniques, which includes self-adaptive tuning, dynamic task scheduling, and resource allocation.

2.5 Conclusion

In this chapter, we walk through a brief introduction of the major failure mechanisms in nanometer CMOS technologies. The physical fundamentals of the failure mechanisms are presented, as well as the most popular physical models associated with them. While there are still open issues related to the physical fundamentals of those models that need further clarification efforts, there is a clear lack of compact models, which are highly desirable for reliability simulation and assessment. Next, we present existing work of aging sensors on reliability characterization and assessment. We also discuss the shortcoming of the existing designs utilization in a proactive DRM system and conclude that, as a key component for quantitative reliability management, novel aging sensors dedicated to online degradation assessment are required. Subsequently, we present state of the art reliability aware computation techniques tuned for combating aging degradation. Those proposals can be fitted into four categories, namely aging-resistant architecture design, aging-aware synthesis, self-adaptive tuning, and dynamic task scheduling and resource allocation.

As we have already discussed in this chapter, a proactive DRM system has to

be constructed upon the solid understanding of the underlying physical process of aging failure mechanisms. In the next chapter, we present a unified NBTI and HCI aging model for FinFET devices, based on the RD model. This accurate aging model constitutes the foundation stone of our proactive reliability-aware computing platform.

3

Unified Aging Model for Dynamic Reliability Management

As planar MOSFET is approaching its physical scaling limits, FinFET becomes one of the most promising alternative structure to keep on the industry scaling-down trend for future technology generations of 22 nm and beyond. In this chapter, we propose a unified reliability model of Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) for double-gate and triple-gate FinFETs, towards a practical reliability assessment method for future FinFETs based circuits. The model is based on the Reaction-Diffusion (RD) theory and it is extended to cover the FinFET specific geometrical structure. Apart of introducing the reliability model we also investigate the circuit performance degradation due to NBTI and HCI in order to create the premises for its utilization for assessing and monitoring the aging process of FinFET based Integrated Circuits (ICs). To validate our model we simulate NBTI and HCI degradation and compare the obtained V_{th} shift prediction with the one extracted out of experimental data. The simulation results suggest that our model characterize the NBTI and HCI processes with accuracy and it is computationally efficient, which makes it suitable for utilization in reliability-aware architectures as reliability prediction/assessment kernel for lifetime reliability management.

3.1 Introduction

As CMOS device scaling entered the deep sub-micron regime, the combination of extremely small devices with a supply voltage that cannot scale proportionally leads to a rising reliability concern due to multiple degradation mechanisms, such as NBTI, HCI, TDDB, and so on [23, 38, 62, 82, 95]. Moreover,

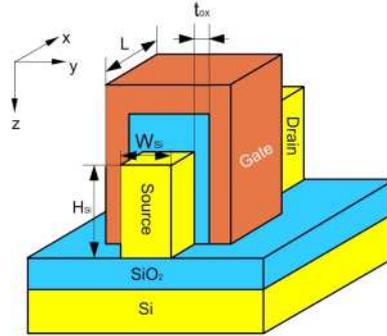


Figure 3.1: SOI FinFET Schematic View.

with the commence of the nanoscale electronics era, worries about reliability are even worsened, as the bottom-up manufacturing processes are inherently prone to defects and the nano fabricated IC runtime faults are expected to be high, caused by shrunken device dimensions and low supply voltages. Moreover, as the conventional MOSFET planar structure is approaching its physical limits, the emerge of novel nanoscale devices raises further reliability concerns because of the introduced new materials and device structures.

Among various emerging devices for future nanotechnology circuits and systems, the multi-gate field effect transistors, e.g., MuGFETs, FinFET [24, 92], represents one of the most promising alternative candidates to replace planar CMOS devices, due to its improved electrostatic controllability and drive current, therefore reduced Short Channel Effect (SCE) relative to the bulk CMOS technology. However, the new geometrical features of these new structures introduce new degradation processes to the device.

In this chapter, we propose a unified NBTI and HCI degradation model for double-gate and triple-gate FinFET devices. The model falls under the RD theory [14, 55, 59] framework and captures the FinFET specific geometrical aspects. The major contributions of the chapter can be summarized as follows:

- A unified NBTI and HCI degradation model for double-gate and triple-gate FinFET devices is proposed. The proposed model unifies the NBTI and HCI degradations, which simplifies the simulation complexity thus makes it suitable for utilization in circuit simulation.
- The reduced dimension effects of nanoscale devices on the NBTI and HCI induced degradations are thoroughly investigated.
- Moreover, based on this reliability model, we introduce a device perfor-

mance degradation model able to capture and predict the aging process due to NBTI and HCI inside FinFET based ICs.

To validate our model we simulate the NBTI and HCI degradations and compared the obtained V_{th} shift prediction with the one extracted out of experimental data. The simulation results indicate that our model characterizes the NBTI and HCI process with accuracy and it is computationally efficient, which makes it potentially applicable for lifetime reliability management schemes to be included in reliability-aware architectures.

3.2 FinFET Device

The FinFET device, also known as "multigate" device, refers to a MOSFET which incorporates more than one gate into a single device. The idea of FinFET device was first introduced by a UC Berkeley research team in 1999 [43]. Compared with planar devices, the FinFET multiple-gate structure provides a better channel controllability. As a result, the FinFET device has a superior suppression of the undesirable short-channel-effect. The leakage current during the "off-state" is decreased and the "on" state drive current is enhanced. These advantages translate into lower power consumption and better performance. Based on these advantages, the ITRS has predicted that FinFET devices will be the cornerstone of sub-32 nm technologies [3].

Despite all the advantages that FinFET can have, their novel geometrical structure raises new reliability concerns. Recent experimental investigations indicate that MuGFET devices with standard orientation are more susceptible to NBTI than planar devices due to the higher availability of Si-H bonds at the (110) oriented fin sidewalls [38]. Furthermore, a self-heating effect [106] caused by the SOI body may speed up the NBTI degradation for its thermal-activated nature. FinFET exhibits an improved HCI immunity with decreasing fin width [23], however, its immunity significantly depends on several factors, such as interface state generation, self heating effect, and temperature-dependent bandgap energy [64]. As a result, NBTI and HCI remain major reliability concerns for the FinFET devices and circuits.

3.3 Generalized RD Model and Its 1-D Solutions

Within all the degradations a device may experience during its operational life, the NBTI and HCI stresses have a very similar physical progress. Both of

these degradations are related to the generation of interface traps causing Si dangling bonds at the Si/SiO₂ interface. NBTI is prominent in PMOS devices along the entire channel when negative gate-to-source voltage is applied, while HCI is prominent in NMOS devices and occurs near the drain end due to the "hot" carriers accelerated in the channel. The interface traps accumulate at the Si/SiO₂ interface then cause a threshold voltage (V_{th}) shift, which results in a poor drive current and shorten device and circuit lifetime.

NBTI and HCI has been well studied and understood for the planar structures as MOSFETs. Traditionally, NBTI degradation is modelled by the Reaction-Diffusion (RD) model [14, 55, 59], and HCI degradations by the "lucky-electron" model [40] and then is shifted to an energy-driven model [91]. As suggested in [61], HCI can also be modelled under RD theory framework, and a geometry-dependent unified RD model for NBTI and HCI has been proposed for planar and surround-gate MOSFETs [55]. However, that model doesn't capture all the FinFET's features, e.g., the high aspect (H_{Si}/W_{Si}) ratio. Moreover, the impact of NBTI and HCI degradations on the circuit performance has not been well studied yet. A general and accurate reliability model is essential for circuit designs and for lifetime reliability management frameworks. Since both degradations can be modelled under the same RD framework, a unified aging model of NBTI and HCI stresses provides benefits, e.g., simplicity and efficiency, for Dynamic Reliability Management (DRM) implementations.

3.3.1 Generalized RD Modeling Framework

As mentioned in previous section, NBTI and HCI are physically induced by the Si-H/Si-O bonds breakage at the Si/SiO₂ interface. NBTI occurs when the gate node is negatively biased and it is accelerated at a elevated temperature condition. Holes from the inversion layer can tunnel into the gate oxide, break the Si-H bond leaving behind an interface trap. While in the HCI process, the energetic "hot" electrons are accelerated by the lateral electric field in the channel and can be injected into the oxide near the drain end causing interface traps too. The H atoms released in this process diffuse away from the Si/SiO₂ interface. Consequently, interface charges are induced, raising the device threshold voltage V_{th} .

Conventionally, NBTI and HCI are modelled separately. While it might be possible to evaluate these degradations individually in experiments and testing, it is hard and of practical relevance to separate them in real operating circuits. Therefore, when targeting an online IC lifetime management, a unified aging model is of interest as it reduces the resource overhead and the complexity of

the monitoring process.

Given the initial concentration of the Si-H bonds, i.e., N_0 , and hydrogen species profile function $H(x, t)$ (H stands for atomic hydrogen particles H^0 and H^+ , or hydrogen molecule H_2), the kinetic equation describing the interface reaction is

$$\frac{\partial N_{it}(0,t)}{\partial t} = k_f(N_0 - N_{it}) - k_r N_{it} H(0, t)^{1/a}, \quad (3.1)$$

where N_{it} is the interface state concentration, and k_f and k_r are the forward and reverse reaction rates, respectively. $H(0, t)$ is the time-dependent hydrogen density at the Si/SiO₂ interface, while a is the kinetic exponent (1 for atomic hydrogen particles H^0 and H^+ , and 2 for hydrogen molecule H_2). It is worth to mention that the forward reaction rate k_f has a dependence on the vertical electrical field E_{ox} .

The generated hydrogen species diffuse away from the interface towards the gate, driven by the gradient of the density. The process is governed by the Fick's Second Law [73], as follows

$$\frac{\partial H(x, t)}{\partial t} = D_H \cdot \nabla^2 H(x, t) \pm q_H \cdot \mu_H E_{ox} \frac{\partial H}{\partial t}, \quad (3.2)$$

where D_H and μ_H are the diffusion coefficient and mobility of species H respectively, E_{ox} is the electric field in the oxide, and q_H is the charge state of species H ($q = 0$ for H^0 and H_2 , and $q = 1$ for H^+). This diffusion equation is geometry dependent, thus its solution requires the fabrication technology details of the device under consideration, i.e., FinFET in our case.

In weak electric-fields, D_H and μ_H are field-independent and follow the Einstein's relation

$$\frac{\mu_H}{D_H} = \frac{q}{kT}, \quad (3.3)$$

where q , k , and T are unit electron charge, Boltzmann's constant, and absolute temperature, respectively.

And the number of new generated interface traps is given by the total number of mobile hydrogen atoms, which can be expressed as

$$N_{it}(t) = a \int_0^\infty H(x, t) dx. \quad (3.4)$$

Eq. (3.1) and Eq. (3.2) defines the basic framework of RD theory. Specifically, Eq. (3.1) is the reaction equation, which defines the generation of the interface

traps; and Eq. (3.2) is the diffusion equation, which defines the concentration profile of the hydrogen species in the gate oxide. The solutions of these equations will be discussed in the next section.

3.3.2 Solutions of 1-D Reaction-Diffusion Equation

A complete solution for the reaction-diffusion process defined by Eq. (3.1) is given in [79], which gives a very complex analytical expression for $N_{it}(t)$ and $H(x, t)$. Another frequently referred solution for Eq. (3.1) is based on a triangular approximation of hydrogen species profile along x-axis. In this approximation, the hydrogen species are assumed to be neutral (i.e., H is H^0 and/or H_2), which is suggested by Alam et al. in their work [14, 55]. Applying the triangular approximation of H species profile to Eq. (3.4) yields

$$\begin{aligned}
 N_{it}(t) &= a \int_0^{x(t)} H(x, t) dx \\
 &= a \int_0^{\sqrt{D_H t}} H(0, t) \left(1 - \frac{x}{\sqrt{D_H t}}\right) dx \\
 &= \frac{a}{2} H(0, t) \sqrt{D_H t},
 \end{aligned} \tag{3.5}$$

where $x(t) = \sqrt{D_H t}$ is the approximated diffusion front, and $H(0, t)$ is the particle concentration at the Si-SiO₂ interface.

According to experimental measurements, the interface trap generation rate is very slow [13]. Hence, it is safe to assume that $\partial N_{it}/\partial t \approx 0$. Moreover, we can also assume that the initial interface trap density $N_{it}(0)$ is negligible when compared to the density of Si-H bond, i.e., $N_{it}(0) \ll N_0$. Thus, Eq. (3.1) can be simplified as

$$N_{it} H(0, t)^{1/a} \approx \frac{k_f}{k_r} N_0. \tag{3.6}$$

By substituting Eq. (3.5) into Eq. (3.6), we can obtain the final expression for $N_{it}(t)$. For atomic hydrogen we have

$$N_{it, H_0}(t) = \sqrt{\frac{k_f N_0}{2k_r}} (D_H t)^{1/4},$$

while for molecule hydrogen we have

$$N_{it, H_2}(t) = \left(\frac{k_f N_0}{2k_r}\right)^{2/3} (D_{H_2} t)^{1/6}.$$

For the positive charged hydrogen atoms (i.e., protons), a rectangular approximation can be assumed [39], which gives

$$N_{it,H^+}(t) = \sqrt{\frac{k_f N_0}{2k_r}} (\mu_{H^+} E_{ox} t)^{1/2}.$$

Eq. (3.7) to Eq. (3.7) suggest a time evolution exponent of 1/4, 1/6, and 1/2 for NBTI-induced interface trap generation with H^0 , H_2 , and H^+ as diffusion hydrogen particles, respectively. In experimental results [7, 54, 83, 86], the 1/6 time evolution exponent is usually found, thus the hydrogen molecule is preferable during diffusion.

3.4 FinFET Reduced Dimension Effect on NBTI

All the discussions thus far are based on the assumption that the oxide dielectric is thick enough to ignore the gate-electrode effect, thus the diffusion of hydrogen species is considered to be in a semi-infinite space with a rectangular plane source located at $x=0$. This assumption can hardly hold for current and future advanced technology nodes ($t_{ox} \leq 2nm$). However, the exact effect of gate electrode (either reflecting or absorbing hydrogen particles) remains unknown yet, thus the semi-infinite diffusion assumption is still adopted here. On the other hand, as the device dimension scaling continues, the rectangular cross section of hydrogen diffusion assumption has to be modified, especially for the novel device geometric structure cases like FinFET.

3.4.1 Diffusion Source Limited Size Effect

In [54], the effect of the geometry on hydrogen diffusion is covered by introducing 2-D and 3-D diffusion for the line-edge and corner effect, respectively. The authors divided the diffusion space into rectangular, cylindrical, and spherical sections, which requires a "regular" shape of cross section. In the following we present a more generalized way to model the geometric effect.

Instead of separating the diffusion space into several independent parts, we propose a quasi 2-D diffusion model to handle the reduced cross-section dimension effect. Figure 3.2 demonstrates the reduced cross section effect in the Triple-Gate (TG) FinFET (Figure 3.2a) and the Double-Gate (DG) FinFET (Figure 3.2b) device, respectively. Starting with the TG-FinFET device, we assume a uniform hydrogen concentration for the gradient line with a distance

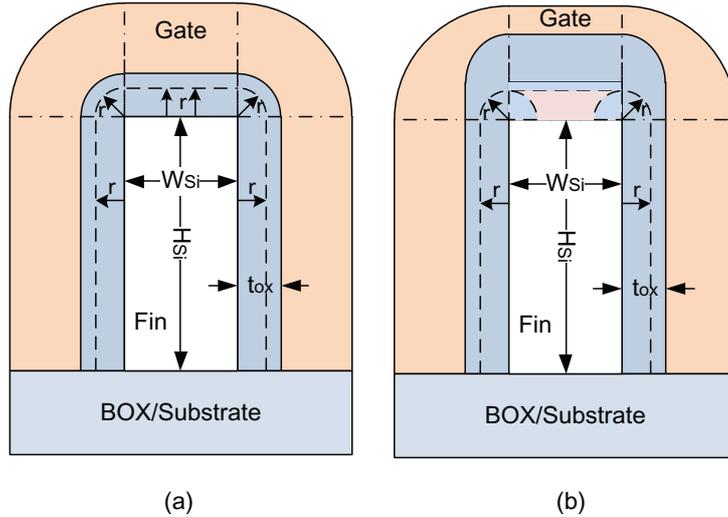


Figure 3.2: Quasi 2-D cross section for hydrogen diffusion: (a) Triple-Gate (TG) FinFET; (b) Double-Gate (DG) FinFET

r respective to the Si-SiO₂ interface, as indicated by the dash-line in Figure 3.2, which leads to a 1.5-D diffusion by transforming the line-edge effect at the top corners into an extended diffusion front line problem.

In the generalized RD framework, Eq. (3.4) defines the geometric dependence. In three-dimensional coordinates, a more general form of Eq. (3.4) can be written as:

$$N_{it}(t) = \frac{a}{S_0} \int_{-\infty}^{\infty} X(V, t) dV \quad (3.7)$$

$$= \frac{a}{S_0} \int_{-\infty}^{\infty} X(r, t) \cdot S(r) dr, \quad (3.8)$$

where S_0 is the area of the diffusion source, $X(V, t)$ is the hydrogen profile at time t , and Eq. (3.8) is a simplified version of Eq. (3.7) if the flux direction of hydrogen particles is always vertical to the surface S . We now introduce a general geometric parameter $G(R^n, t)$ as

$$\begin{aligned} G(R^n, t) &\equiv \frac{N_{it}(t)}{aX(0, t)} \\ &\equiv \frac{\int_{-\infty}^{\infty} X(r, t) \cdot S(r) dr}{X(0, t)S_0}, \end{aligned} \quad (3.9)$$

where R^n means n -dimensional real-space. $G(R^n, t)$ can be considered as a

geometry-dependence coefficient, which defines the diffusion behavior in a specific geometrical structure. We note that this coefficient is time-dependent because the boundary conditions (i.e., the effect of gate-electrodes) are not clearly defined, thus the real geometrical structure of the diffusion is determined by the diffusion front, which is time-dependent.

Given the relation defined by Eq. (3.9), we can then simply give another relation for N_{it} , which is

$$N_{it}(t) = aG(R^n, t)X(0, t). \quad (3.10)$$

Substituting Eq. (3.10) into Eq. (3.6) yields

$$N_{it}(t) = \begin{cases} \sqrt{k_f/k_r N_0 G(R^n, t)} & a = 1 \\ \sqrt[3]{k_f/k_r N_0^3/2 G(R^n, t)} & a = 2 \end{cases} \quad (3.11)$$

When applying the quasi 2-D diffusion assumption to the TG-FinFET, as graphically depicted in Figure 3.2, the $N_{it}(t)$ is expressed as

$$N_{it}^{TG}(t) = \frac{aL}{2H_{si} + W_{si}} \int_0^{\sqrt{D_H t}} X(r, t) \cdot S(r) dr,$$

with $S(r)$, the diffusion front surface, being

$$S(r) = L \cdot (2H_{si} + W_{si} + 2 \cdot 2\pi r/4),$$

where H_{si} , W_{si} , and L are the dimensions of TG-FinFET as indicated in Figure 3.1. By defining the diffusion length $\lambda = \sqrt{D_H t}$, we can obtain the geometry-dependence coefficient for TG-FinFET as

$$G_{TG}(W_{si}, H_{si}, t) = \frac{\lambda}{2} \cdot \frac{2H_{si} + W_{si} + \pi\lambda/3}{2H_{si} + W_{si}}. \quad (3.12)$$

Applying the same procedure to the DG-FinFET structure, we can obtain the geometry-dependence coefficient as follows:

$$G_{DG}(W_{si}, H_{si}, t) = \frac{\lambda}{2} \cdot \frac{H_{si} + \pi\lambda/3}{H_{si}}. \quad (3.13)$$

Assuming that all the released hydrogen is converted into molecule hydrogen, and submitting Eq. (3.12), we obtain the analytical expression of NBTI for TG-FinFET:

$$N_{it}(t) = \left(\frac{k_f N_0}{k_r} \right)^{2/3} \left(\lambda \cdot \frac{2H_{si} + W_{si} + \pi\lambda/3}{2H_{si} + W_{si}} \right)^{1/3}. \quad (3.14)$$

From Eq. (3.14) we can deduce that: when $\lambda \ll (2H_{si} + W_{si})$, we have $N_{it}(t) \approx \lambda/2$ and we obtain a $\sim t^{1/6}$ time-dependence for hydrogen molecule diffusion; when $\lambda \sim (2H_{si} + W_{si})$, we have $N_{it}(t) \sim (1 + \pi/3) \cdot \lambda/2$, which means that the geometry structure will approximately lead to a $\sqrt[3]{2}$ prefactor to hydrogen molecule diffusion; if enough stress time t is given to let $\lambda \gg (2H_{si} + W_{si})$ (i.e., the problem becomes a total 2-D diffusion problem), our model predicts that the time-dependence of $N_{it}(t)$ changes to $\sim t^{1/3}$, which requires further experimental confirmation.

Eq. (3.13) indicates that the DG-FinFET has larger geometric dependence than the TG-FinFET because the effective Si-SiO₂ area in the DG-FinFET is smaller than the one of the TG-FinFET, thus it experiences severer corner effect than the TG-FinFET.

We note that in Eq. (3.12) and Eq. (3.13), G_{TG} doesn't depend on the oxide dielectric thickness. This is a consequence of the fact that because we assumed infinite thickness for it, thus the effective diffusion range is only determined by the diffusion front. It is worth noticing that the root of the effect of dimension degradation defined by Eq. (3.9) is the dimension degradation of the source plane, i.e., the area of Si-SiO₂. In other words, as the dimension of device is scaling down, the corner effect (i.e., the effect of line edge of the diffusion) becomes more and more important, leading to a quasi-2D diffusion of the hydrogen species. Under some extreme conditions (e.g., quantum dot device), the diffusion source can degrade into a quasi 0 dimension source, which will further change the diffusion into a quasi-3D problem.

3.4.2 Finite-Oxide Thickness Effect and Oxide-Gate Interface Role

So far our discussion is based on a "infinite oxide thickness" assumption, which can be hardly considered to be true in advanced technology nodes. If we assume a finite thickness for the oxide dielectric, we have to deal with an unavoidable problem: what's the role of the oxide-gate interface in the diffusion process?

While we have no accurate answer to that question so far, the effect of oxide-gate interface can be one or a combination of: absorbing, reflecting and/or transmitting. Absorbing and reflecting prevents hydrogen particles to diffuse into gate electrode, in either case the oxide-gate interface acts as a barrier to diffusion. Absorbing seems unlikely to happen in real device, as it implies that the hydrogen particles have to vanish at the interface; reflecting will make

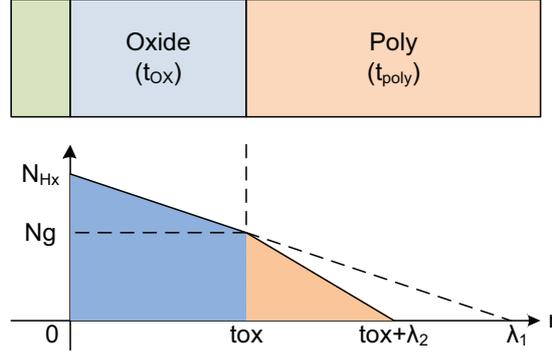


Figure 3.3: Hydrogen concentration profile in finite-size oxide: t_{ox} is the oxide thickness, and gate thickness is considered to be infinite; $\lambda_1 = \sqrt{D_{ox}t}$ and $\lambda_2 = \sqrt{D_{si}t}$.

the triangular approximation for hydrogen profile invalid. In order to simplify our discussion, we adopt the transmitting effect assumption as the role of the oxide-gate interface in the finite-oxide thickness diffusion process.

As indicated in Figure 3.3, we consider a continuous hydrogen concentration at both sides of the oxide-gate interface. According to Eq. (3.11), the N_{it} can be expressed as

$$N_{it}(t) = \frac{a}{S_0} \left(\int_0^{t_{ox}} H_1(r, t) S(r) dr + \int_{t_{ox}}^{t_{ox} + \lambda} H_2(r, t) S(r) dr \right),$$

where $H_1(r, t)$ and $H_2(r, t)$ are the hydrogen concentration function in oxide and gate, respectively. Using the triangular approximation we can get:

$$\begin{aligned} H_1(r, t) &= H(0, t) \left(1 - \frac{r}{\lambda_1}\right) & (0 \leq r \leq t_{ox}), \\ H_2(r, t) &= H(0, t) \left(1 - \frac{t_{ox}}{\lambda_1}\right) \left(1 - \frac{r - t_{ox}}{\lambda_2}\right) & (r > t_{ox}). \end{aligned}$$

By submitting all the equations above into Eq. (3.9), we obtain the geometry dependence coefficient as

$$\begin{aligned} G_{TG,2} &= C(A, t_{ox}, \lambda_1) \\ &\quad + \frac{\lambda_2}{A\lambda_1} (\lambda_1 - t_{ox}) \left[(A + \pi t_{ox}) + \frac{\pi \lambda_2}{6} \right], \end{aligned} \quad (3.15)$$

where $A = 2H_{si} + W_{si}$ and $C(A, t_{ox}, \lambda_1)$ is

$$C(A, t_{ox}, \lambda_1) = t_{ox} + t_{ox}^2 \left(\frac{\pi}{2A} - \frac{1}{\lambda_1} - \frac{\pi t_{ox}}{3\lambda_1 A} \right). \quad (3.16)$$

Given that $t_{ox} \leq \lambda_1$, Eq. (3.16) can be estimated as $C \approx (t_{ox} + \pi t_{ox}^2/2A) - kt_{ox} \approx const$, so the characteristics of Eq. (3.15) are mainly determined by the rest terms of its right-hand-side (RHS). Based on this approximation Eq. (3.15) can be written as

$$G'_{TG,2} = \sqrt{\frac{D_{si}}{D_{ox}}} (\lambda_1 - t_{ox}) \left[\frac{A + \pi t_{ox}}{A} + \frac{\pi \lambda_2}{6A} \right]. \quad (3.17)$$

Eq. (3.17) suggests a similar time evolution dependence with Eq. (3.12): if the term $A + \pi t_{ox}$ is large enough ($(A + \pi t_{ox}) \gg \pi \lambda_2/6$), then the time evolution exponent is determined by $(\lambda_1 - t_{ox})$ and it's $\sim t^{1/6}$ for molecule hydrogen; when $(A + \pi t_{ox}) \ll \pi \lambda_2/6A$, the time evolution exponent becomes $\sim t^{1/3}$ for molecule hydrogen.

In the discussion above, we have presented the general RD theory framework for NBTI failure mechanism in the triple-gate and double-gate FinFET devices, with the consideration of the reduced dimension effect induced by the special geometry structures of those devices. In the next section, we extend our work to accommodate the HCI failure mechanism in FinFET devices as well in order to obtain a "unified" degradation model for both NBTI and HCI failure mechanisms.

3.5 HCI Modeling Using RD Model

HCI is a critical reliability concern, particularly when large electric fields exist in the transistor at operating conditions. The physical mechanism behind HCI relates to the fact that channel carriers gain significant energy from channel electric field and inject themselves into the oxide dielectric at the "pinch-off" point, breaking Si-H bonds at the Si-SiO₂ interface and generating interface traps. Given the similarity between NBTI and HCI on interface trap generation, it might be possible to use RD framework to model the parameter shift caused by HCI as well. Actually, in [55] an attempt has been made to make use of the RD model to explain the time-evolution exponent for both NBTI and HCI. Even though the authors succeeded in contributing the $t^{1/2}$ characteristic of HCI to the 2-D hydrogen diffusion, the derivation doesn't capture the entire picture of N_{it} generation under HCI stress. Details related to the 2-D diffusion formatting under HCI and the link between the RD model and the more commonly used energy-driven model [91] are not presented and should be given in order to have a complete unified model. Moreover, of special interest in the geometry-dependence of reliability for nanoscale device, the reduced

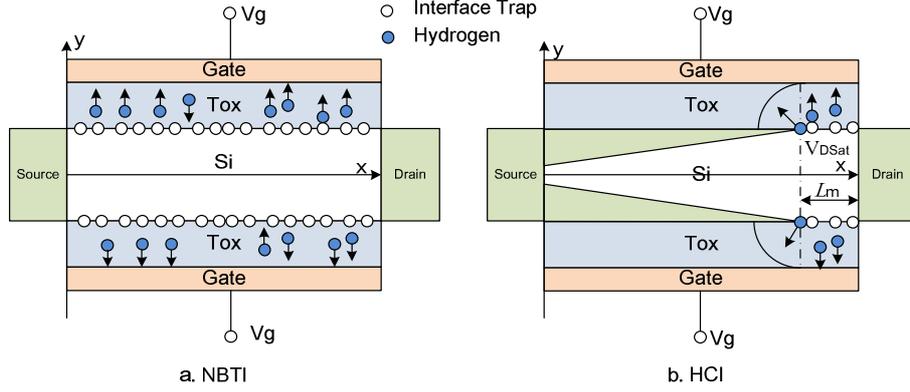


Figure 3.4: Interface traps generation and hydrogen diffusion in DG-FinFET Channel (top-view): (a) NBTI in DG-FinFET; (b) HCI in DG-FinFET.

dimension effect on HCI stress should be taken into consideration as well.

Figure 3.4 illustrates the interface-trap generation and hydrogen diffusion due to NBTI and HCI in a DG-FinFET. The difference between NBTI and HCI is that interface traps generate in a region from pinch-off point to the drain (i.e., velocity saturation region with a length of L_m as indicated in Figure 3.4(b)) under HCI stress, while interface traps generate along the entire channel surface under NBTI stress. In [55], the authors assumed that the length of velocity saturation region L_m is small enough such that the source of hydrogen diffusion can be considered as a line source, which is resulting in a 2-D hydrogen diffusion at the drain. This assumption is quite arguable since according to [17] L_m can be estimated as

$$L_m = \sqrt{\frac{2\epsilon_{Si}}{qN_A} [V_{DS} + \phi_{bi} - (V_{DSSat} + \phi_0)]}, \quad (3.18)$$

where N_A is the channel dopant concentration, ϕ_0 is the bulk potential, and ϕ_{bi} is the build-in p-n junction potential. Assuming that $\phi_0 \approx \phi_{bi}$, we obtain

$$L_m = \sqrt{\frac{2\epsilon_{Si}}{qN_A} (V_{DS} - V_{DSSat})}, \quad (3.19)$$

which is not negligible for state of the art devices. In fact, the diffusion source under HCI stress still has to be considered as a reduced 2-D surface and requires the quasi-2D RD framework we presented in the previous section.

If we assume a uniform distribution of N_{it} along L_m and triangular concentration approximation for hydrogen, the interface traps generated by HCI stress

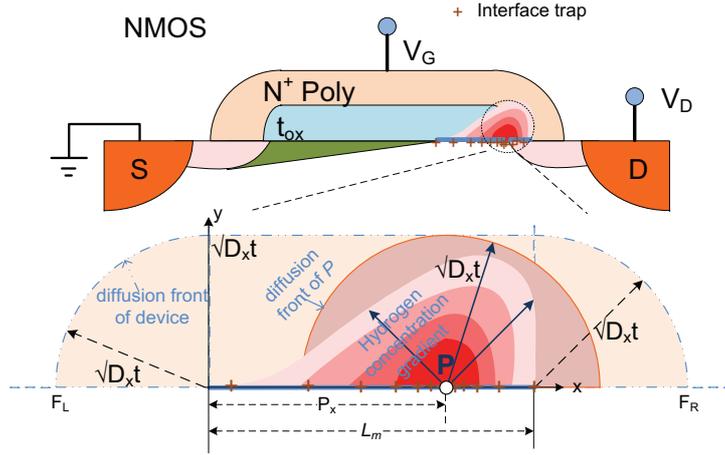


Figure 3.5: Detailed view of interface traps generation and hydrogen diffusion in Planar Structure MOSFET: P is the location of worst case HCI stress, F_L and F_R are the left and right diffusion front on x axis, respectively.

then can be expressed as:

$$\begin{aligned}
 N_{it,HCI}(t) &= a \int_0^\lambda H(r, t) S(r) dr \\
 &= a \int_0^\lambda H\left(1 - \frac{r}{\lambda}\right) \left(L_m + 2 \cdot \frac{\pi r}{2}\right) dr \\
 &= \frac{a\lambda H(0, t)}{2} \left(L_m + \frac{\pi\lambda}{6}\right),
 \end{aligned}$$

where $\lambda = \sqrt{D_X t}$. According to the definition of geometry-dependence coefficient in Eq. (3.9), we obtain

$$G_{HCI}(t) = \frac{\lambda}{2} \left(L_m + \frac{\pi\lambda}{6}\right). \quad (3.20)$$

Assuming that all the released hydrogen particles are converted into hydrogen molecules and submitting Eq. (3.20) into Eq. (3.11), we obtain the following analytical expression of N_{it} for HCI

$$N_{it,HCI}(t) = \left(\frac{k_f N_0}{k_r}\right)^{2/3} \left(\frac{\lambda}{2} \left(L_m + \frac{\pi\lambda}{6}\right)\right)^{1/3}. \quad (3.21)$$

This equation indicates that the time exponent n of the HCI degradation is in a range of $(1/3, 2/3)$ depending on the pinch-off region length. The frequently

reported time exponent $t^{0.5}$ falls in this region. Unfortunately, $N_{it,HCI}(t)$ in Eq. (3.21) does not explicitly depend on any geometric parameter, because the "infinite-oxide" assumption has already ruled out the oxide thickness influence.

To link the RD model to the energy-driven model, we have to consider the following details about the damage caused by HCI stress: the generation of interface traps caused by HCI stress is most likely non-uniformly distributed in the L_m region, i.e., $N_{it} \sim N_{it}(x, t)$, because the potential distribution in the channel is non-uniform, thus the forward reaction rate k_f is non-uniform along the surface. This non-uniform distribution of N_{it} further induces an irregular shape of hydrogen diffusion front surface. Without loss of generality, this phenomenon can be illustrated as depicted in Figure 3.5 using a planar structure MOSFET. If we assume that N_{it} follows a distribution function $f(x)$, which is defined as

$$f(x) = \partial N_{it}(x, t) / \partial x \quad (0 \leq x \leq L_m), \quad (3.22)$$

there will be a point "P" which has $df(x)/dx = 0$ that defines the worst case of HCI stress. The parameter shift caused by HCI is determined by the location of this point (P_x in Figure 3.5). Since the interface traps generated here can be considered to be a line source, thus the diffusion of hydrogen is a completely 2-D diffusion. To obtain the exact location of P, we need to solve the non-linear RD equations along with the condition in Eq. (3.22), which requires significant deriving effort and is a little bit out of the scope of this dissertation. In fact, the P_x can be estimated as follows by introducing a fitting parameter k_p that

$$P_x = k_p \cdot L_m. \quad (3.23)$$

k_p is technology dependent and should be obtained by testing after fabrication. Unlike NBTI failure mechanism, HCI does not have a direct dependence on the device geometry structure. In other words, the reduced dimension effect in nanoscale devices has less influence on the HCI failure mechanism.

With the discussions thus far, the physical processes, i.e., the interface trap generation, of NBTI and HCI failure mechanisms have been unified under the same RD theory framework. The utilization of the unified aging model in the lifetime reliability management is presented in the next section.

3.6 Model Utilization in Lifetime Reliability Management

In practice, devices experience multiple concurrent failure mechanisms during the normal operating conditions. With the unified aging model presented above, the amalgamated parameter degradation induced by NBTI and HCI can be obtained and be utilized in lifetime reliability management. The details of utilizing the unified model in lifetime reliability management are presented in the rest of this section.

3.6.1 Lifetime and Aging Definition

To an electronic device/circuit, the term "lifetime" means the time until an important material/device parameter degrades to a point that the device/circuit can no longer function properly in its intended application. However, the damages (i.e., device shifts) induced by NBTI and HCI are gradual in time, and the definition of "function properly" is arbitrary. Hence, by defining ΔV_{th} reaching some critical fraction P_{crt} (in practice, this critical portion of a parameter degradation is defined to be 10% [73]) of the voltage threshold to be the "end-of-life", the aging status and its upper limit can be expressed as

$$\frac{\Delta V_{th}}{V_{th}} \leq P_{crt}, \quad (3.24)$$

where the threshold voltage shift induced by NBTI and HCI, ΔV_{th} , can be calculated by the expression

$$\Delta V_{th} = \frac{q\Delta N_{it}}{C_{ox}}, \quad (3.25)$$

where q is the unit electron charge, ΔN_{it} is the number of interface traps, and C_{ox} is the oxide capacitance.

3.6.2 Degradation Under Random Stress

In order to implement an online reliability management system, the current reliability status must to be first understood. The model we introduced in the previous section is assumed to work under DC conditions. NBTI degradation has a well-known recovery sub-process, by which the NBTI induced damage can partially recover when the stress is released [14].

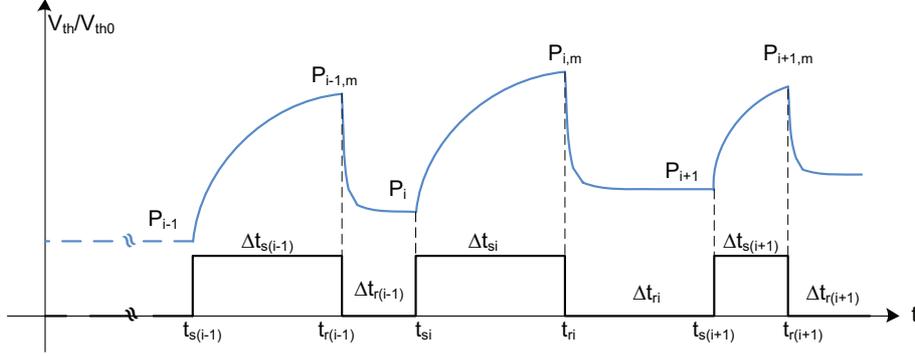


Figure 3.6: NBTI Under Random Stress

Defining $P(t) = \Delta V_{th}(t)/V_{th0}$, where V_{th0} is the threshold voltage of fresh device. And the threshold voltage shift ΔV_{th} then can be expressed as $V_{th0} \cdot \Delta P(t)$. Figure 3.6 illustrates the recovery effect of NBTI under random stress. As one can find in Figure 3.6, the $P(t)$ under random stress can be expressed as

$$\Delta P(t_n) = \sum_{i=1}^n \Delta P_i(t_i), \quad (3.26)$$

in which we divide the entire stress experience into n periods $P_i(t_i)$, and in each and every period the device experience a stress time Δt_{s_i} and recovery time Δt_{r_i} . The maximum point in every period is denoted as $P_{i,m}$ and occurs at the end of the stress sub-sequence. Let $D_i(t)$ and $R_i(t)$ be the degradation and recovery function in i^{th} period, respectively, ΔP_i is

$$\Delta P(t_i) = D_i(\Delta t_{s_i}) - R_i(\Delta t_{r_i}). \quad (3.27)$$

Given that $D_i(t)$ has a form $D_i(t) = At^n$, and $R_i(t)$ has a form $R_i(t) = R_0\sqrt{\xi t}$ ([57]), $\Delta P_i(t_i)$ can be expressed as

$$\Delta P(t_i) = A(\Delta t_{s_i})^n(1 - \sqrt{\xi \Delta t_{r_i}}). \quad (3.28)$$

Substituting Eq. (3.28) into Eq. (3.29), we can obtain

$$\Delta P(t_1, t_n) = \sum_{i=1}^n A(\Delta t_{s_i})^n(1 - \sqrt{\xi \Delta t_{r_i}}). \quad (3.29)$$

Further if we assume that $\Delta t_{s_i} = s_i \Delta t_c$ and $\Delta t_{r_i} = r_i \Delta t_c$, Eq. (3.29) can be

expressed as

$$\Delta P(t_1, t_n) = A\Delta t_c^n \cdot \sum_{i=1}^n s_i(1 - \sqrt{\xi' r_i}), \quad (3.30)$$

where Δt_c is the clock period of the system, and $\xi' = \xi \Delta t_c$. From Eq. (3.30) we can see that the degradation of the system is determined by the stress-recovery pair (i.e., $s_i \sim r_i$), no matter the order of their occurrence and frequency. Using Eq. (3.30), we can obtain the degradation expected value of $\Delta P(t_1, t_n)$ under a specific stress-recovery pattern.

3.7 Results and Discussion

To verify our proposal, NBTI and HCI V_{th} shift prediction are evaluated based on experimental data in [38], [64], and [52] and compared with the degradations predicted by our model. Since in this chapter we focus on the prediction of the evolution of the degradations in time, we carry on the model evaluation on an inverter.

The main parameters of our model are as follows: (i) the reaction rate constants, k_f and k_r , and the diffusion front distance D_H , all of them being temperature dependent. For these parameters, the Arrhenius model [6] is utilized then all of them turn to depend on an activation energy, E_f , E_r , and E_H , respectively. Those activation energies can be measured by experiments and only depend on materials; (ii) the technology and geometry related parameters, e.g., ϵ_{Si} , N_0 , N_A , H_{Si} , W_{Si} can be derived from technology specifications; (iii) electronic parameters, which are V_{DS} , V_{DSSat} , E_m , L_m , can be calculated making use of SPICE alike models; and (iv) the fitting parameters introduced in our model. There is a modulation fitting parameter k_p in Eq. (3.23). Normally, the worst case HCI stress location is very close to the pinch-off point, i.e., $P_x \approx L_m$, thus k_p can be estimated as 1 and the reduced dimension effect can be ignored for HCI.

The results are depicted in Figure 3.7 for NBTI and in Figure 3.8 for HCI. Note that in the figures the solid lines correspond to the predicted values while the individual points correspond to values computed based on experimental data. One can observe that our proposal provides a good accuracy when compared with data derived from experiments. Worth to mention that for NBTI we obtain a root-mean-square-error less than 2%. The HCI prediction is less accurate however, since HCI has a dependence on the lateral electric field (E_m) and gate voltage, while NBTI just depends on the gate voltage only. In order to

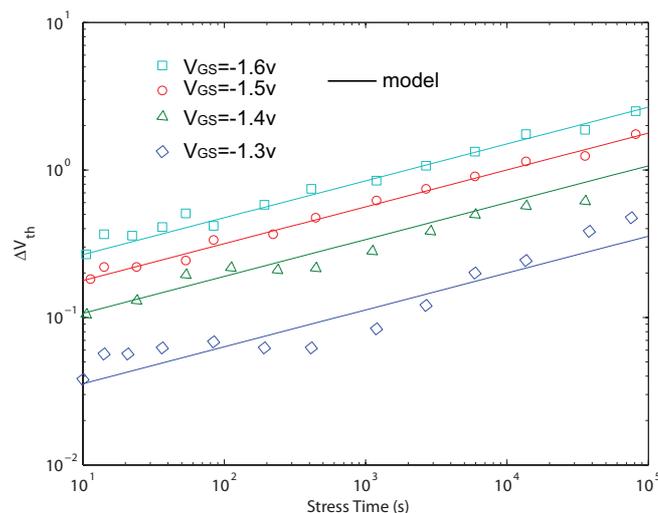


Figure 3.7: V_{th} Degradation Due to NBTI.

improve the accuracy of HCI model, more precise model for the lateral electric field distribution and pinch-off length should be introduced and this constitute a future work subject.

By modeling NBTI and HCI stresses in a unified RD model framework, the number of fitting parameters in our model decrease significantly, then it further simplifies the parameter extraction model computation progress. Hence, a higher performance can be achieved by our model when comparing with other models, e.g., [112]. Furthermore, the degradation caused by NBTI and HCI stresses is indicated by a single proxy, i.e., the threshold voltage shift ΔV_{th} . This induces a reduced resource overhead in a reliability management framework implementation, since only one kind of aging sensor for the threshold voltage is required. In view of that the proposed model can be implemented with less resources and requires less computation effort, thus it enables a potential simple monitoring and prediction agent for reliability-aware computation architectures and platforms.

We also conducted a simple evaluation of the proposed unified degradation model, which results are depicted in Figure 3.9. In the simulation, the threshold voltage shift ΔV_{th} is calculated by using the proposed unified aging model with both NBTI and HCI failure mechanisms under periodic signal stress (duty cycle = 50%). From Figure 3.9, it can be observed that the absolute V_{th} shift increases fast initially and change rate decreases after that. This is in confor-

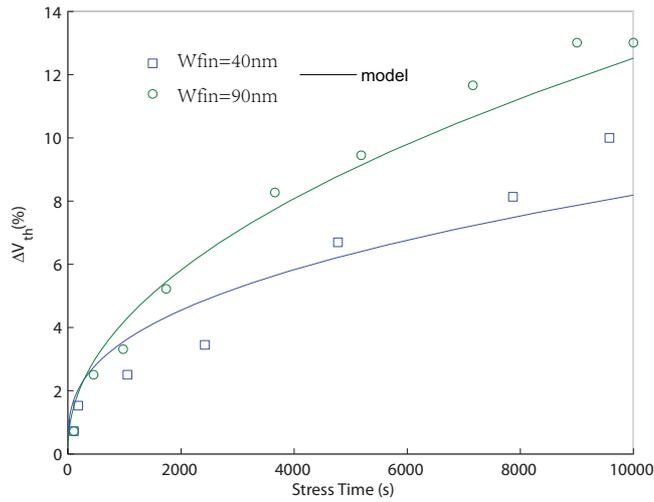


Figure 3.8: V_{th} Degradation Due to HCI.

mity with the power-law of the hydrogen diffusion rate, therefore the interface state generation rate.

3.8 Conclusion

In this chapter we proposed a unified reliability model for Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) degradation specific for double-gate and triple-gate FinFETs and circuits. The proposed model is based on the reaction-diffusion theory and it is extended such that it covers the FinFET specifics. We also investigated the circuit performance degradation due to NBTI and HCI in order to create the premises for the utilization of our proposal in assessing and monitoring the Integrated Circuits (ICs) aging process. To validate our model we simulated NBTI and HCI degradation and compared the obtained V_{th} shift prediction with the one evaluated based on experimental data. Our simulations suggest that the proposed model characterize the NBTI and HCI process with accuracy and it is computationally efficient, therefore it is appropriated for the implementation in lifetime reliability management to make assessment and prediction on the device/circuit degradations.

With the degradation model proposed in this chapter, we can assess and/or

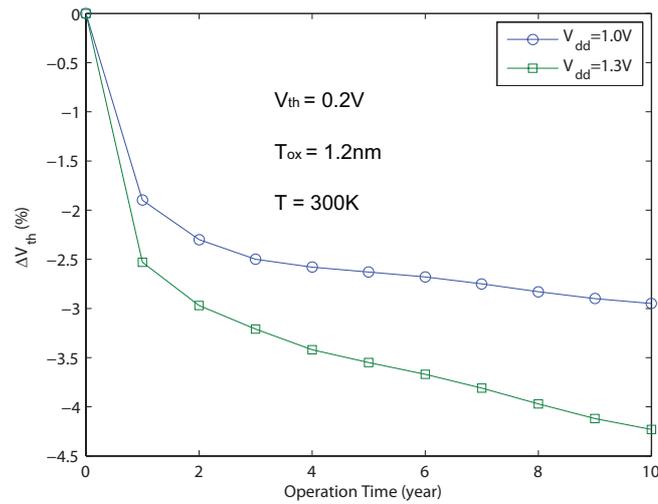


Figure 3.9: V_{th} Degradation Due to the NBTI and HCI Combined Effect.

predict the reliability status of the device/circuit by extracting a physical proxy parameter (e.g., threshold voltage V_{th}) from it. In order to achieve this goal, in the next chapter, we will present two types of aging sensor designs to perform such kinds of parameter measurements.

Note. The contents of this chapter is based on the the following papers:

Y. Wang, S. D. Cotofana, L. Fang, A Unified Aging Model of NBTI and HCI Degradation towards Lifetime Reliability Management for Nanoscale MOSFET Circuits, Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), 2011, pp. 175–180.

4

Aging Sensor Designs for Dynamic Reliability Management

Accurate and efficient degradation measurements are highly desirable for Dynamic Reliability Management (DRM) systems. In this chapter, we propose two types of aging sensor designs, based on threshold voltage (V_{th}) and power supply current (I_{DD}) measurement, respectively, to assess reliability status of devices/circuits. The V_{th} -based aging sensor is highly tolerant to process, voltage, and temperature variations, which is highly desirable for accurate reliability assessment. While the I_{DD} -based aging sensor can extract the amalgamated aging effects from a large block of circuit, which can significantly reduce the required number of aging sensors for probing degradation information in a large system. With degradation information measured from aging sensors, quantitative reliability-aware resource management is made possible. Both sensor designs are verified by simulations in Cadence tools using TSMC 65 nm technology library. The simulation results for the V_{th} -based aging sensor design indicate that the sensor has a very low process, supply voltage, and temperature (PVT) sensitivity, which outperforms the accuracy of state of the art NBTI and HCI sensors under PVT-variation circumstances. The simulation results for the I_{DD} -based aging sensor design indicate that the power supply current exhibits a similar aging rate as the threshold voltage for the entire circuit lifetime, but with a better sensitivity towards the End-of-Life (EOL), which demonstrates the validity and practical relevance of the proposed I_{DD} based aging monitoring framework.

4.1 Introduction

Usually, Integrated Circuits (ICs) lifetime requirements are mostly determined based on worst-case assumptions, which leads to highly conservative margins on technology parameters, resulting in the under utilization of the fabrication technology potential. To make better use of the available technological improvement, the pessimistic assumptions should be relaxed and combined with a Dynamic Reliability Management (DRM) framework that relies on online sensors to measure the ICs aging status. The DRM concept was first introduced in [97] in an attempt to boost performance within an acceptable reliability margin. The idea was then extended in, e.g., [19], which introduced task scheduling and Dynamic Voltage/Frequency Scaling (DVFS) into the DRM system. The influence of process and temperature variation in DRM is considered for the first time in the work [119]. However, most of these proposals do not embed mechanisms to get the actual aging status of the hardware platform, thus they end up with a "blind" optimization for reliability policy. Recently, a DRM system with an aging sensor designed to detect delay violations was proposed in [11]. The central point behind this proposal is a stability checker able to detect delay violation caused by circuit aging effects. Any instability appearing in a "guard-band" is considered to be an aging failure, and then self-adaptive procedures are launched in order to adjust the system's configuration. However the proposed aging sensor is built on a binary detection of timing violation in the critical path, thus its efficiency and adaptivity are limited.

In the recent past, a number of approaches for aging/reliability monitoring have been reported. In [53], Kim et al. introduced an on-chip aging monitor for high resolution NBTI degradation measurements by detecting the beat-frequency of a pair of ring oscillators. Keane et al. further extended this idea to an "all-in-one" sensor for BTI, HCI, and TDDB degradation measurement [44]. Though high precision can be achieved by their circuitry, a large area overhead is required (0.035mm^2 in 130nm technology). Karl et al. proposed compact in-situ sensors for monitoring NBTI and TDDB, respectively, in [48]. These sensors work in the sub-threshold region to increase the sensitivity. Even though they require a small area overhead, they are sensitive to process, voltage, and temperature variations. Agarwal et al. proposed aging sensor designs to be integrated inside flip-flops to detect delay violation(s) in [10, 11]. These designs are relatively small and can be potentially included in many chip flip-flops. However, this kind of sensor can only check delay violation in a static or quasi-static time window ("guardband"), thus they cannot provide quantitative aging information.

Given that existing DRM approaches have different limitations, especially they don't consider the non-uniform dynamic aging progress on chip, we introduce in this chapter a novel DRM framework for reliability-aware applications tailored for current and future technologies. The proposed DRM utilize online aging sensor circuitry to measure the NBTI/HCI induced degradation. These on-chip aging sensors monitor either the transistor threshold voltages or the circuit power supply current, which carry on a direct indication of the aging severity in the circuit. Subsequently, the collected aging profiling data are further processed by a reliability assessment module to determine the current reliability state, which is a key information for reliability management. Furthermore, based on the same data, a reliability prediction module gives a forecast on "future" reliability state of the system, which can be utilized to provide failure warnings and serves as a more indicative information for reliability optimization. The major contributions of the chapter can be summarized as follows:

- A new DRM framework that relies on quantitative degradation measurements is introduced. This employs new aging sensors for NBTI/HCI induced degradation, which are monitoring the V_{th}/I_{DD} deviations caused by dynamic environmental stresses. Different than aforementioned prior works, our approach can quantitatively analyze the system aging status, which gives a deeper insight on the degradation progress and eventually leads to a better optimization for reliability.
- Novel V_{th} based aging sensors for NBTI and HCI degradation measurement are proposed with good Process, Voltage, and Temperature (PVT) variation tolerance. Though, thus far, substantial work have been done on designing different types of aging sensors and circuits, most of them are targeting the characterization of failure mechanisms. On the contrary, our work focus on designing aging sensor for lifetime reliability managements, thus to achieve appropriate performance and accuracy, the environmental variations are considered in our designs.
- A Novel I_{DD} based aging sensor is proposed with the capability of directly measuring the actual IC aging status caused by the amalgamated effects of NBTI and HCI failure mechanisms.

The rest of the chapter is organized as follows: Section 4.2 presents the V_{th} based aging sensors and DRM system. Section 4.3 describes the circuit design of V_{th} sensors for NBTI and HCI. Section 4.4 presents the V_{th} sensors performance evaluation and a comparison with the state of the art. Section 4.5

presents the I_{DD} based aging sensor and DRM system. Section 4.6 describes the circuit design of I_{DD} sensor and its components. Section 4.7 presents the I_{DD} sensor performance evaluation. Section 4.8 concludes this chapter.

4.2 V_{th} Based Aging Sensors and DRM System

Among all the failure mechanisms, NBTI stress on PMOS transistors have been considered to be a dominant limiting factor of device's lifetime. NBTI is prominent in PMOS devices along the entire channel when negative gate-to-source voltage is applied. It causes a positive shift in the absolute value of threshold voltage ($|V_{th}|$) with time [111], thereby resulting in poor drive current and circuit delay degradation. Traditionally, HCI stress is considered to be less critical than NBTI stress, but with decreasing gate thickness and length, and increasing channel electrical field, HCI is becoming a major reliability concern also. Moreover, variations from fabrication process technology, as well as dynamic environment like supply voltage and temperature, cause a statistical degradation across circuits and chips, thus securing ICs' lifetime reliability at design time is becoming more difficult than ever. Consequently, introducing a Dynamic Reliability Management (DRM) scheme to assure IC's lifetime specifications in advanced process nodes becomes necessary.

Figure 4.1 presents an overview the proposed DRM system framework with V_{th} based aging sensors. As depicted in the figure, the aging sensors are attached to the output signals of the Combinational Logic Blocks (CLBs). Each aging sensor takes one output signal from CLB as the input stress signal. The proposed aging sensors can measure the device threshold voltage and convert it into a *Delay* signal, which is further converted into the digital domain by an asynchronous counter. The converted delay signal is then recorded into the aging profile database and can be further processed by DRM software. Generally, the aging sensor is such designed that the *Delay* signal is proportional to the interested physical parameter (i.e., V_{th} in our case), thus the recorded values of *Delay* signals are direct reflections of the reliability status in the devices/circuits. The *Monitor* and *Pulse* signals are used to control the working mode of aging sensors.

The V_{th} based aging sensor is the key component of the aging sensor, as it extracts up to date aging information from the device under observation. When the sensor enters the measuring mode, a *Pulse* signal is sent to the Voltage Controlled Delay Line (VCDL), which generates a delayed *Pulse* signal during the measurement. The time delay added by the VCDL is proportional to the

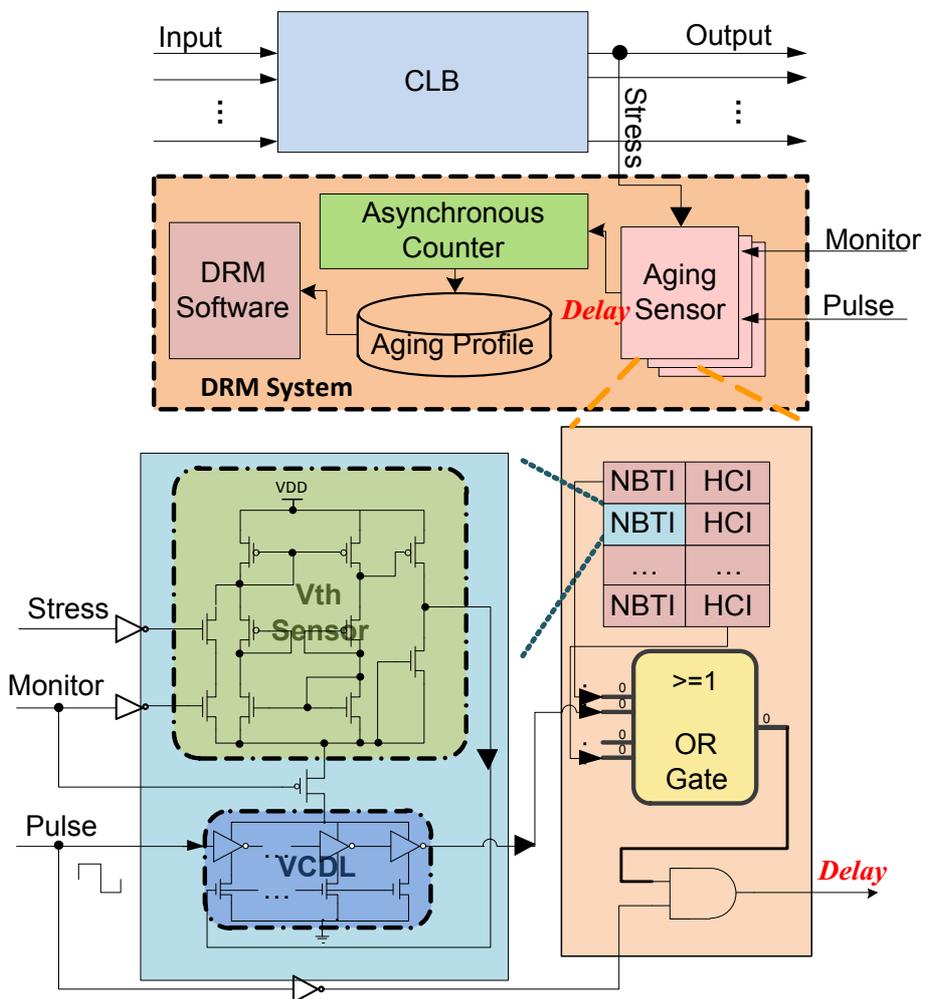


Figure 4.1: Schematic of the Proposed DRM System (the Upper Part) and Illustration of the V_{th} Based Aging Sensors (the Lower Part).

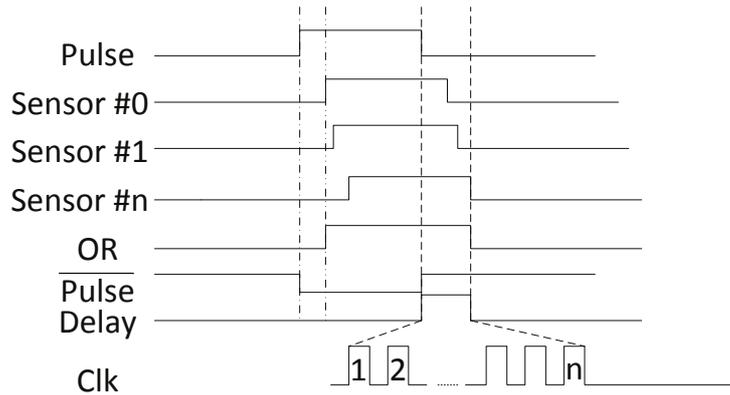


Figure 4.2: Signal Waveform for Degenerated Delay Measurement of Aging Sensors.

input V_{th} signal from the aging sensor. By comparing the original and delayed *Pulse* signals, a *Delay* signal which represents the V_{th} shift can be obtained.

In order to reduce the area consumption, measurements from multiple sensors are fused into one output signal using an OR gate as indicated in Figure 4.1. In such a way, the most aged path inside the same logic block is then selected. This idea is based on the assumption that the combinational logic inside a same function unit experience similar environmental conditions, e.g., temperature, activity ratio, and voltage variation. This assumption is generally true based on the correlation of hardware in the same function unit. Violations on this assumption could be eliminated by a careful selection of sensor's placement. In [114] an algorithm to find critical aging path was presented and can be utilized in our system as well.

Figure 4.2 depicts the signal waveform for the delay measurement of the aging sensors. A *Pulse* signal triggers the VCDL in the aging sensor, and the output of the V_{th} sensor is used as the VCDL control signal. The VCDL propagation delay is proportional to the V_{th} signal magnitude, which directly reflects the aging status of the logic under observation. The *OR* signal presents the degenerated signal for multiple sensors. The degenerated *OR* signal is further *AND* with an inverted *Pulse* signal as indicated by the \overline{Pulse} signal in the figure. Finally, the outputted *Delay* signal is extracted and sent to an asynchronous counter and could be digitized using the system clock signal.

Note that variations exist commonly in a chip in the deep sub-micron technologies, thus a certain number of sensors have to be embedded into the chip to capture the statistical nature of degradation, which puts severe constraints on sensor area and power consumption. On the other hand, the sensor circuitry

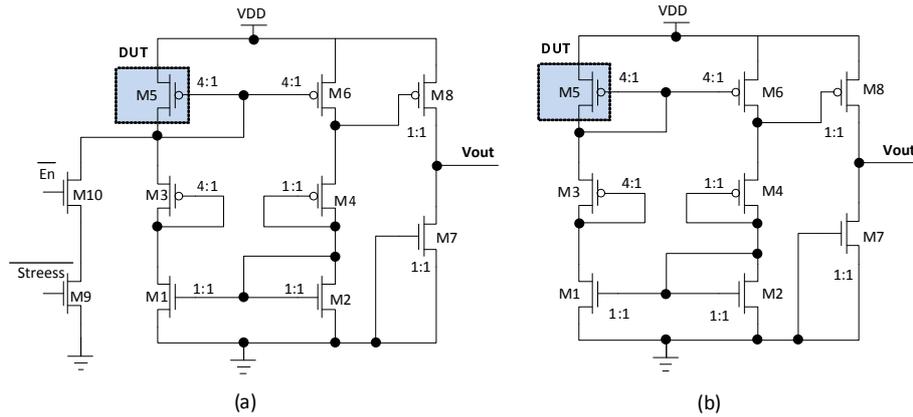


Figure 4.3: (a) NBTI Sensor Circuit Schematic; (b) Measuring Mode Equivalent Circuit Schematic.

should be insensitive to the environment variations to provide a precise measurement. In the next section, the detailed circuit design of the V_{th} based aging sensors for NBTI and HCI are presented with the consideration of variation resistance.

4.3 Circuit Design of V_{th} Sensors

Figure 4.3(a) depicts the proposed sensor circuit able to detect long-term V_{th} degradation due to NBTI, where M5 is the Device Under Test (DUT). The basic idea of this design is to make use of the V_{th} extractor circuit from [116] to monitor the M5 V_{th} value. Transistors M9, M10, and M11 are used to control the working mode of the sensor, where MS is the control signal. When MS is set to "1" and the enable signal \overline{En} is set to "0" the sensor works in measurement mode. M9 is turned off and M11 is turned on, then the \overline{Stress} signal through M9 is detached from the sensor and a current path through M11 is open, allowing the V_{th} extractor circuit formed by M1 ~ M8 to work. The equivalent circuit of the NBTI sensor during measurement mode is presented in Figure 4.3(b). On the opposite, if MS is set to "0", the sensor works in stress recording mode. M11 is turned off and no current can flow in the V_{th} extractor. All the transistors in V_{th} extractor circuit are turned off except transistor M5, which is attached to a discharging path controlled by the \overline{Stress} signal through M10.

To extract the V_{th} from M5, transistors M1, M2, M5, and M6 are con-

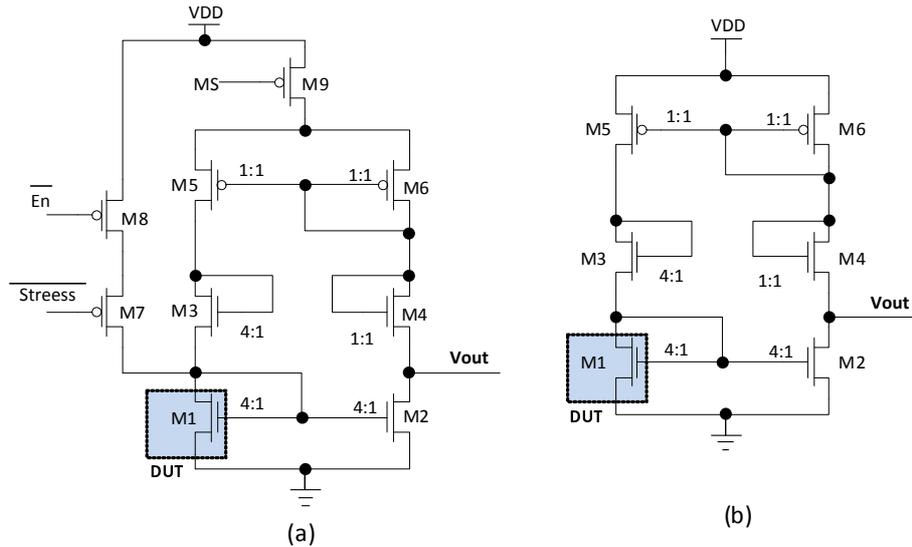


Figure 4.4: (a) HCI Sensor Circuit Schematic; (b) Measuring Mode Equivalent Circuit Schematic.

nected head to tail to form a closed-loop current mirror (also known as "beta-multiplier current reference"), such that the currents flowing through these transistors are forced to be equal by the feed back loop. This circuit structure has a good tolerance on PVT variations, and it is portable to different technologies, which increase the accuracy and flexibility of the sensor design. The size ratio of transistors in the V_{th} extractor indicated in Figure 4.3 the voltage at the terminal " V_{out} " is the $M5$ absolute V_{th} value.

Figure 4.4(a) depicts an HCI sensor circuit using a beta-multiplier current reference similar with the proposed NBTI sensor above. In the HCI sensor, M1 is the DUT, and transistors M7, M8, and M9 are used to control the working mode of the sensor, where MS is the control signal. When MS is set to "0" and enable signal En is set to "1", the sensor works in measuring mode. M7 is turned off and M9 is turned on, then the \overline{Stress} signal through M8 is detached from the sensor, allowing the V_{th} extractor circuit formed by M1~M6 to work. On the opposite, if MS is set to "0", then M9 is turned off and no current can flow in the V_{th} extractor. All the transistors in the V_{th} extractor are turned off, except transistor M1, which is attached to a pull-up path controlled by the \overline{Stress} signal through M8. Since M7 is turned on, the input voltage of M1 then is controlled by the \overline{Stress} signal. When \overline{Stress} is "1", M8 is turned off, then the gate voltage is pulled down to GND through M1 and M1 turns off; if when

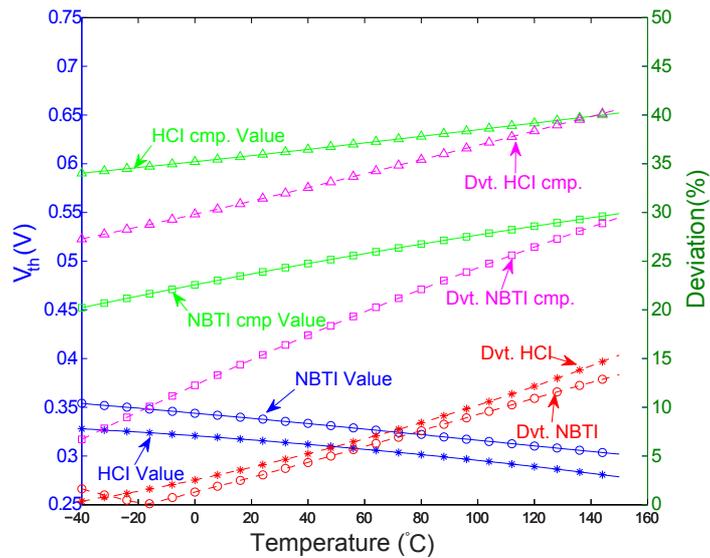
\overline{Stress} is "0" then M8 is on, pulling the input of M1 to be V_{DD} , so M1 is turned on. In other words, M1 experiences the same switching activity as the stress signal, so M1 degrades due to HCI stress.

4.4 V_{th} Sensors Evaluation

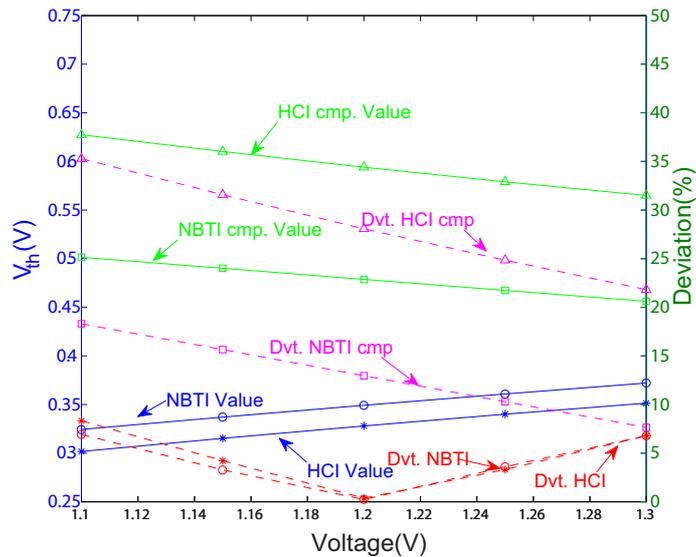
To evaluate PVT-variation influence on the proposed circuits, simulations are run using the TSMC 65 nm technology library ($V_{DD} = 1.2$ V). The temperature variation range is set to be -40 °C to 150 °C and the voltage variation is 1.1 V to 1.3 V, which is about 10% deviation from the standard V_{DD} value. The process variation are simulated using a Monte-Carlo simulation in the Cadence Virtuoso environment. For comparison purpose, we performed the same simulations for the aging sensor introduced in [51]. A normal V_{th} value under normal conditions ($T = 27$ °C and $V_{DD} = 1.2$ V) is extracted from each sensor's implementation using Cadence Virtuoso tools and these values are used as reference.

Figure 4.5(a) presents the dependence of the NBTI/HCI sensor V_{th} measurements to temperature variation: the left axis is the V_{th} value extracted from the sensors, and the right axis is the absolute V_{th} deviation relative to the nominal V_{th} value of the technology extracted at the normal conditions ($T = 27$ °C and $V_{DD} = 1.2$ V), and the data with "cmp" label are for the sensors introduced in [51]. We can observe that the output voltage of the proposed sensors varies from about 280 mV to 350 mV, as temperature increases from -40 °C to 150 °C, with a temperature sensitivity of around 0.29 mV/°C. The deviation is maximum 15% for extreme conditions and about 5% at room temperature range. The temperature sensitivity of the design from [51] is about 0.51 mV/°C for the NBTI sensor and 0.325 mV/°C for the HCI sensor, with a deviation ranges from 8% to 30% and from 30% to 40% for the NBTI and HCI sensors, respectively.

Figure 4.5(b) presents the dependence of the NBTI/HCI sensor V_{th} measurements to supply voltage variation: the left axis is the V_{th} value extracted from the sensors, and the right axis is the absolute V_{th} deviation relative to the nominal V_{th} value of the technology extracted at the normal conditions ($T = 27$ °C and $V_{DD} = 1.2$ V). The figure suggests that the output voltage of the proposed sensors have a V_{DD} -variation sensitivity of around 0.24 mV/mV. The deviation is maximum 7% for extreme conditions and negligible around the standard V_{DD} value. As a comparison, the V_{DD} -variation sensitivity is about 0.23 mV/mV for the NBTI sensor and 0.25 mV/mV for the HCI sensor



(a) Temperature Dependence ($V_{DD}=1.2$ V)



(b) V_{DD} Variation Dependence ($T = 27$ °C)

Figure 4.5: Temperature and V_{DD} Variation Dependence: the left axis is the V_{th} value from the sensors, and the right axis is the absolute deviation relative to the normal conditions ($T = 27$ °C and $V_{DD} = 1.2$ V) and the data with "cmp" label are for the sensors introduced in [51].

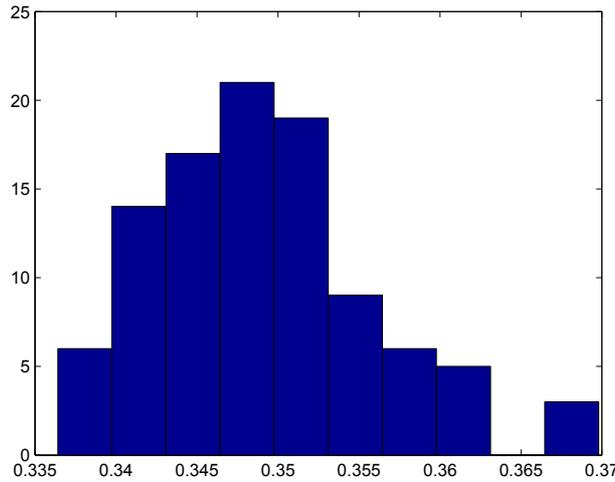
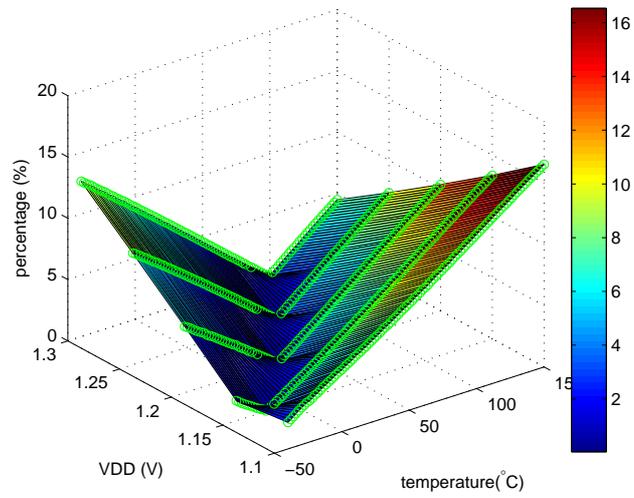


Figure 4.6: Histogram Plot of the Output V_{th} of NBTI Sensor using Monte-Carlo Simulations.

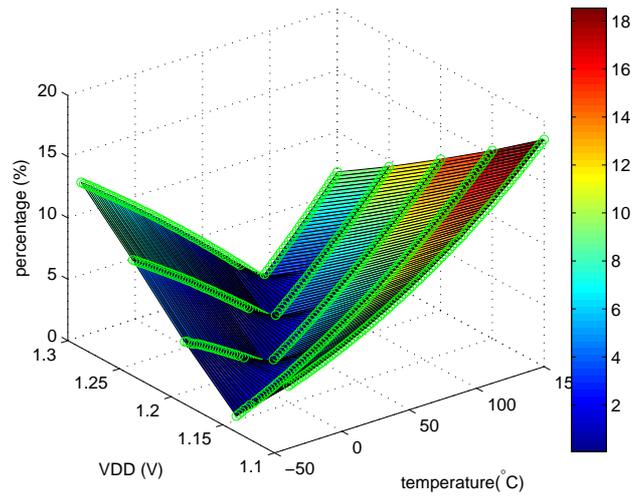
in [51], with a deviation ranges from 8% to 18% and from 22% to 35% for the NBTI and HCI sensors, respectively.

In order to evaluate the process variation tolerance of the proposed sensors, Monte-Carlo simulations (run for 100 times) are conducted and the results are depicted in Figure 4.6. From the histogram plot one can observe that the extracted V_{th} values are concentrated in the range from 0.34 V to 0.355 V and the entire V_{th} value span is less than 35 mV. Therefore, we can safely conclude that the proposed sensors are process mismatch tolerant.

Figure 4.7(a) and Figure 4.7(b) present the deviation of the measured threshold voltage relative to the nominal V_{th} value at operating condition $T = 27\text{ }^{\circ}\text{C}$ and $V_{DD} = 1.2\text{ V}$, in the presence of both temperature and supply voltage variations. As one can observe in the figures, for the extreme conditions, the deviation is about 16% for the NBTI sensor, and 17% for the HCI sensor. If we assume a temperature range ($10\text{ }^{\circ}\text{C}$, $60\text{ }^{\circ}\text{C}$) and $\pm 0.05\text{ V}$ of V_{DD} as being a normal variation range for daily operating environments, the worst case error produced by the sensor is less than 8%. Additionally, as one can deduce from the simulation results, the deviations have a good linearity in both temperature and voltage variations, which simplify the online calibration of the measurement process.



(a) NBTI Sensor



(b) HCI Sensor

Figure 4.7: V_{th} Deviation Relative to Normal Condition ($T = 27^\circ\text{C}$ and $V_{DD} = 1.2\text{ V}$) with Temperature and Voltage Variations.

The area overhead of our design is about $5.25 \mu m^2$, and the power consumption is about $15.39 \mu W$ using TSMC 65 nm technology (for V_{th} sensors only). Furthermore, as opposed to Karl's work [48] which requires a complicated empirical model to calibrate the ΔV_{th} to the output frequency, the relationship between V_{th} value and sensor's output of our design is simple and linear, which gives a very simple calibration scheme for our design. Most importantly, the work mentioned above is focusing on characterizing failure mechanisms, while our work is dedicated to DRM.

To conclude, the proposed V_{th} -based aging sensors are process, voltage, and temperature tolerant and straightly address the requirements of a DRM system at the expense of relatively small area overhead and low power consumption. These features make them a good candidate to be implemented in DRM systems to dynamically assess the reliability status from the devices/circuits.

4.5 I_{DD} Based Aging Sensors and DRM System

While the threshold voltage (V_{th}) is the most common physical parameter selected as indicator of transistor aging progress, the V_{th} -based aging sensors can only monitor the aging status at transistor level. Thus, for circuit level aging measurements many such sensors are required. Moreover, sensor positioning and the extrapolation method that can bring aging information from transistor to circuit level are far from being trivial issues. Furthermore, V_{th} is hard to be extracted directly without interrupting circuits' normal function. As a result, all V_{th} based aging sensors make use of sacrificed devices in order to replicate the stress to which the circuit under observation is subjected to, leading to an indirect measurement. In order to measure the real aging status inside a circuit, we propose to use the power supply current (I_{DD}) as aging indicator.

In Figure 4.8, the block diagrams of the V_{th} -based and I_{DD} -based aging measurement schemes are illustrated. Different from existing V_{th} -based measuring scheme, the I_{DD} -based measuring scheme measures the power supply current directly from the Circuit-Under-Observation (CUO). As depicted in Figure 4.8(b), the proposed I_{DD} -based sensor consists of a Built-In Current Sensor (BICS), which mirrors the transient I_{DD} current of the CUO, and sends it to a Current-mode Peak Detector (CPD). The CPD detects the peak value of the input current by using a current comparator, and holds the peak current for an adjustable time within a current memory, which allows the Current-to-Time converter (C2T) to translate the current value into a Pulse-Width Modulated (PWM) signal. With the PWM signal, the aging status of the CUO can be ex-

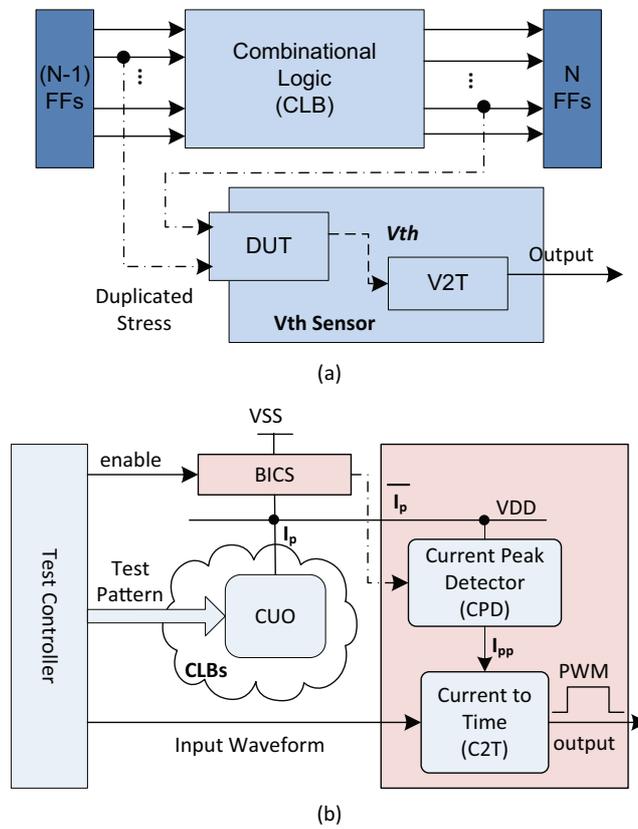


Figure 4.8: Different Measurement Schemes for Degradation Detection: (a) V_{th} sensor scheme; (b) a direct measurement scheme with the proposed I_{pp} sensor (in the right box). The aging indicator I_{pp} of the proposed sensor is taken from the Circuit-Under-Observation (CUO) directly. V_{th} sensor takes the aging indicator V_{th} from the Device-Under-Test (DUT) of the sensor.

tracted by further processing with the model discussed in Section 4.5.1. This aging information can be further utilized to implement a DRM system, which can provide the best system performance for certain given application and reliability requirements.

4.5.1 I_{DD} Degradation Model Due to Aging

Power supply current I_{DD} is the total drain current passing through the supply voltage terminals. Without loss of generality, we utilize the inverter circuit in Figure 4.9(a) as the discussion vehicle to determine the relationship between I_{DD} degradation and failure mechanisms. Figure 4.9(b) depicts the inverter Voltage-Transfer-Curve (VTC). During a low-to-high (i.e., the input voltage switching from "0" to "1") or high-to-low input transition, the inverter working point is determined by the intersections of the output characteristic curves of the PMOS and NMOS transistor, as depicted in Figure 4.9(c). According to the conduction status of the PMOS and NMOS devices, the operating regions of the inverter can be divided into 5 regions, which are graphically illustrated in Figure 4.9(d). During this process, at some point ($V_{in} = V_{out} = V_M$), both transistors are saturated and the power supply current I_{DD} reaches the maximum value I_{pp} ($I_{pp} = \max(I_{pp})$) because the V_{DS} values are equal for both transistors. Further increase of input voltage makes the NMOS enter the linear region and as a consequence the drain current I_D decreases towards 0 (because the PMOS is switched-off by the gate overdrive voltage). The analysis is similar for the high to low input transition (when the input voltage changes from "1" to "0"). Summarizing, the current I_{DD} reaches a peak I_{pp} during any input signal transition, when $V_{in} = V_{out}$.

The above analysis can be easily applied to more complex CMOS networks such as the general CMOS logic structure, depicted in Figure 4.10(a). The PMOS devices are equivalent to a pull-up network and the NMOS devices are equivalent to a pull-down network. Since the input vectors to the pull-up and the pull-down networks are complementary in CMOS logic, the working regions of pull-up and pull-downs networks shift oppositely during input signal transitions. Thus at some point, the I_{DD} of the entire network reaches a peak I_{pp} . For the fresh (unaged) devices, this peak value is constant for a given input pattern, so the degradation of the peak current can be chosen as indicator for assessing the aging status inside any (large) CMOS logic network.

Based on the above analysis and since the global I_{DD} is just a special case of I_D (i.e., when the total drain current from both the pull-up and the pull-down networks reaches its maximum), we can consider the peak power supply

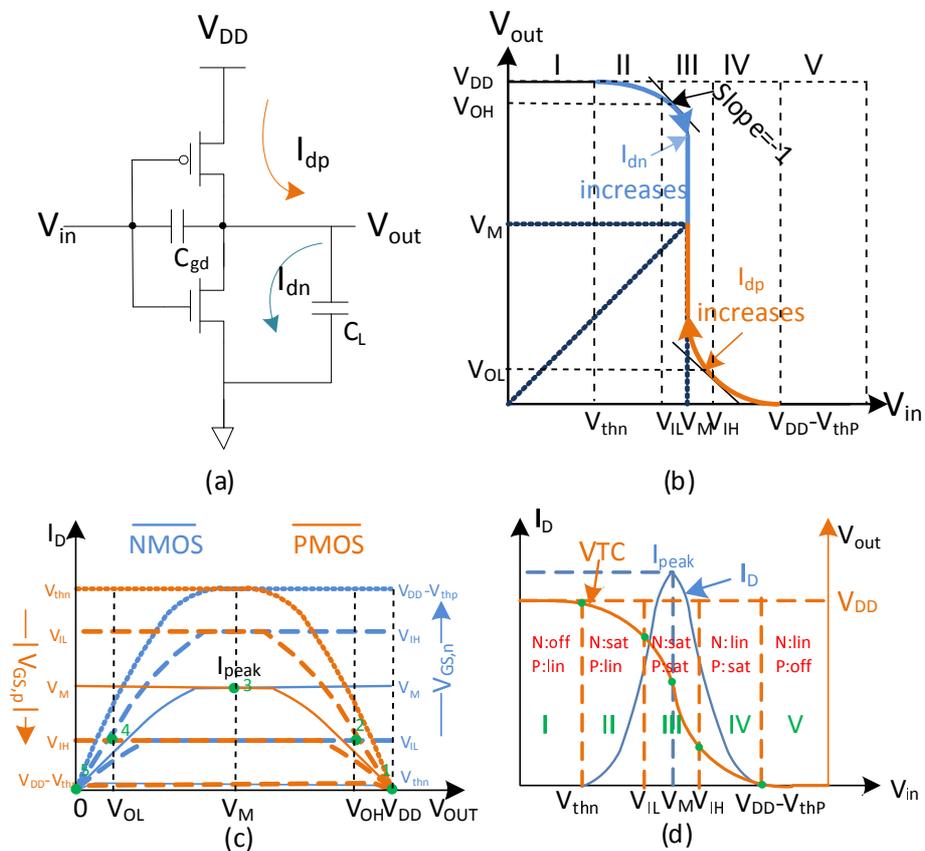


Figure 4.9: Inverter peak current: (a) the circuit model; (b) Voltage-Transfer-Curve (VTC) and the operating regions of PMOS, NMOS transistors; (c) intersections of the output characteristic curves of PMOS, NMOS transistors; (d) the peak power supply current and transistor operating regions.

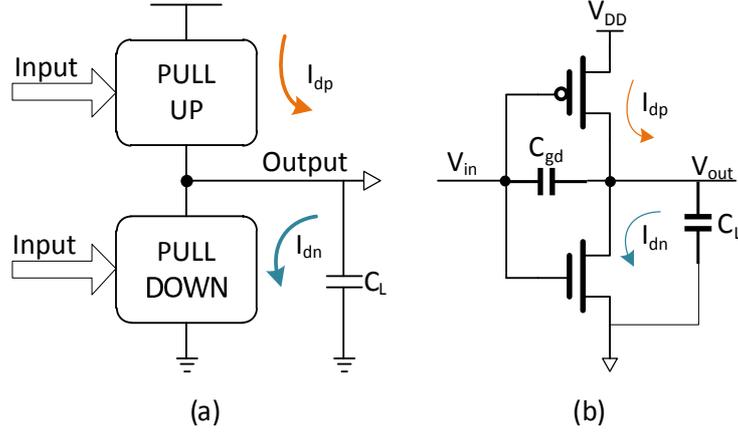


Figure 4.10: Peak current of CMOS logic: (a) a general illustration of CMOS network; (b) Equivalent invert circuit for the pull-up network.

current as a signature of the drain current. Therefore, without loss of generality, we derive the aging model for I_{DD} using the drain current.

Generally, the drain current in the saturation region can be expressed as:

$$I_{D_{sat}} = \mu C_{ox} \frac{W}{L} (V_{GS} - V_{th})^2, \quad (4.1)$$

and a more general expression of the drain current is

$$I_D = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right]. \quad (4.2)$$

Eq. (4.1) and (4.2) indicate the relationship between I_D and key device parameters, such as the carrier mobility μ , the device threshold voltage V_{th} , and the oxide capacitance C_{ox} , which are degrading under joint NBTI and HCI induced stress. The peak power supply current is a special case of I_D , i.e., the I_D value which is equal in both the pull-up and the pull-down network. Therefore, as far as aging effects caused by wearout mechanisms are concerned, I_{pp} follows the same rule as I_D does.

NBTI is an intrinsic front-end-of-line wearout mechanism, which occurs in PMOS transistors mainly when the gate is subjected to a negative input voltage. The NBTI-induced I_D damage includes trap generation at the channel-dielectric interface as well as inside the bulk of the dielectric. Consequently, the threshold voltage V_{th} shifts and channel mobility μ degrades due to the

trap generation. The V_{th} degradation induced by NBTI can be expressed by:

$$\Delta V_{th} = \frac{q\Delta N}{C_{ox}}, \quad (4.3)$$

where q is the electron charge and ΔN is the total trap density generated by NBTI. The mobility degradation can be described by the following equation [25]:

$$\mu = \frac{\mu_0}{1 + \alpha\Delta N}, \quad (4.4)$$

where μ_0 is the original channel mobility, α is a process-dependent constant, and $\alpha \approx 2.4 \times 10^{-12} \text{ cm}^2$ [63]. For small ΔN , using $\Delta\mu = \mu_0 - \mu$, the channel mobility degradation can then be estimated by:

$$\frac{\Delta\mu}{\mu} = \frac{\Delta N}{1 + \alpha\Delta N} \approx \alpha\Delta N. \quad (4.5)$$

Assuming that the threshold voltage shift and the channel mobility degradation are independent progresses, the change of I_D can be expressed as:

$$\Delta I_D = \frac{\partial I_D}{\partial V_{th}} \Delta V_{th} + \frac{\partial I_D}{\partial \mu} \Delta\mu. \quad (4.6)$$

Since the NBTI stress happens when PMOS is in saturation mode, applying the above equation to Eq.(4.1), yields:

$$\Delta I_D = \frac{I_{D0}}{(V_{GS} - V_{th} - V_{DS}/2)} \Delta V_{th} + \frac{I_{D0}}{\mu_0} \Delta\mu. \quad (4.7)$$

Generally, $V_{DS} \approx 0$ when the channel is conducting. Inserting Eq.(4.3) and Eq.(4.5) into the above equation, we can estimate the degradation of I_D due to NBTI as follows:

$$\frac{\Delta I_D}{I_{D0}} = \frac{q\Delta N}{C_{ox}(V_{GS} - V_{th})} + \alpha\Delta N, \quad (4.8)$$

which suggests that the fraction of I_D degradation is proportional to the severity of NBTI wearout.

Similarly, the HCI-induced degradation is also a combined effect of threshold voltage shift and channel mobility degradation, which implies that basically Eq. (4.6) also holds true for HCI degradation. Hence, the amalgamated aging effects of NBTI and HCI can be written as:

$$\frac{\Delta I_D}{I_{D0}} = \left[\frac{q}{C_{ox}(V_{GS} - V_{th})} + \alpha \right] \cdot \Delta N_{nbt+hci}, \quad (4.9)$$

where $N_{nbt+hci}$ is the total trap density generated by NBTI and HCI.

Eq. (4.9) suggests that the I_{DD} degradation fraction is proportional to the amalgamated aging effects induced by NBTI and HCI. In fact, the aging effects of other failure mechanisms, e.g., Time-Dependent-Dielectric-Breakdown (TDDB), can be indicated by the I_{DD} degradation as well, since the consequences induced by the failure mechanisms, e.g., threshold voltage shift, carrier mobility degradation, and so on, are reflected in the I_{DD} degradation. Therefore, even in a more general case, I_{DD} can be utilized as the indicator of the amalgamated aging effects of multiple failure mechanisms.

4.6 Circuit Design of I_{DD} Sensor

The key component of the proposed sensor is the Current-mode Peak Detector (CPD), which is described in detail in the following sub-section. In order to achieve a good accuracy, the Current-to-Time converter (C2T) is carefully designed as well, and details are presented in Section 4.6.2. For the current sensing we make use of the Built-In Current Sensor (BICS) approach proposed in [35].

4.6.1 The Current Peak Detector (CPD)

Figure 4.11 presents the proposed current peak detector, that is composed of: (i) a current memory cell (consists of M_4 to M_7) with an adjustable memory holding time constant, to retain the peak current value, and (ii) a current comparator (consists of M_1 to M_5) to determine if the present supply current value is bigger than the stored peak value, in which case the peak current value in the memory cell is updated. The current memory cell consists of two regulated cascade stages [30], which enables a better output swing and an increased output impedance.

As depicted in Figure 4.11, I_p is the input current of the CPD, I_{pm} is the stored current in the current memory cell, V_{ctrl} the output voltage of the current comparator, and I_{pp} is output current of the CPD. The holding time of the current memory cell is adjusted by I_{bias} .

The current comparator [21], compares the values of I_p and I_{pm} currents. To describe the CPD operation, we distinguish between two functional stages: (i) the mirroring stage, characterized by $I_p > I_{pm}$ (i.e., the CPD input current is bigger than the stored peak value), and (ii) the peak holding stage, character-

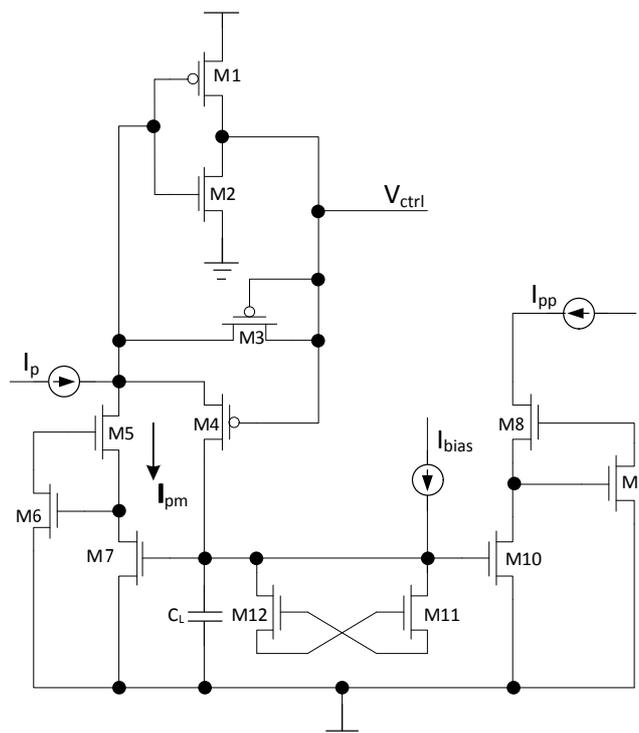


Figure 4.11: Circuit Schematic of the Current-Mode Peak Detector.

ized by $I_p < I_{pm}$ (i.e., the CPD input current is smaller than the stored peak value, hence the peak value remains unchanged).

During the mirroring stage, the comparator input current is positive and the comparator input voltage increases to the upper rail voltage V_{DD} , forcing the output voltage V_{ctrl} to be low, enabling the current memory cell via transistor M_4 , to mirror I_p (i.e., $I_{pp} = I_{pm} = I_p$). During the peak holding stage, V_{ctrl} is high, transistor M_4 is off, and the current memory cell holds the peak value of the I_p current (i.e., $I_{pp} = I_{pm}$). The storing capability is achieved using the gate-to-source capacitance of M_7 and M_{10} , in parallel with the discharge path consisting of cross coupled transistors M_{11} and M_{12} and current source I_{bias} . In this way, the discharging time constant can be controlled by the current I_{bias} . The maximum peak current is limited by the sourcing ability of the comparator with feedback diode, while the minimum detected peak current is constrained by the comparator gain and the output impedance of the circuit under aging assessment.

4.6.2 The Current-to-Time Converter (C2T)

The C2T converter, based on a thyristor delay element [33], is depicted in Figure 4.12. It receives the comparator output voltage V_{ctrl} as triggering input and generates a time interval proportional to the value of the control current I_{pp} . Its operation can be described as follows: When a rising edge of V_{ctrl} is detected, the load capacitance charges from 0 to V_{DD} slowly until the voltage V_{out} reaches the threshold voltage of transistor M_2 and fast for the remaining time to V_{DD} , due to the feedback connection. The falling edge of V_{ctrl} triggers the discharging of the voltage across capacitor C_L through the drain-source capacitance of M_1 and the gate-source capacitance of M_2 . The transient waveform of the voltage across the load capacitance C_L is presented in Figure 4.13.

The C2T time delay is defined as the rising time of the converter voltage output and is given by the the relation:

$$T = C_L \cdot V_{th_{M_2}} / I_{pp},$$

where $V_{th_{M_2}}$ denotes the threshold voltage of transistor M_2 .

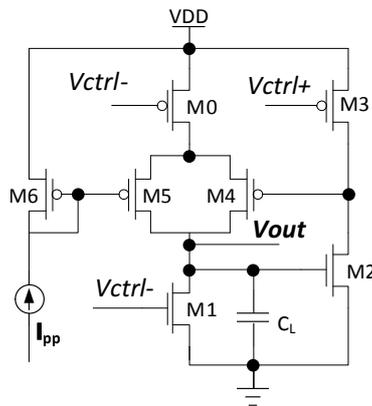


Figure 4.12: Circuit Schematic of the Current-to-Time Converter.

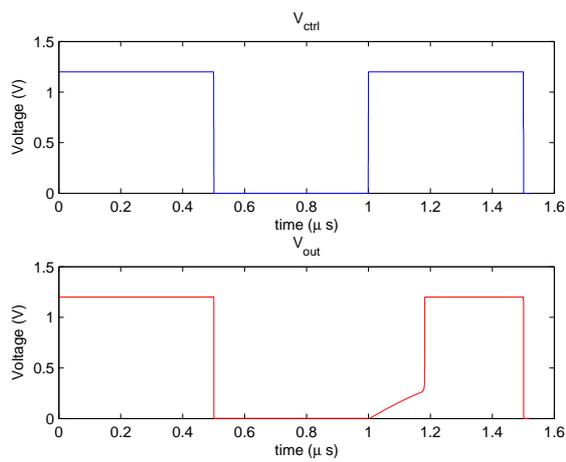


Figure 4.13: The Transient Waveform of the Voltage across Capacitor C_L .

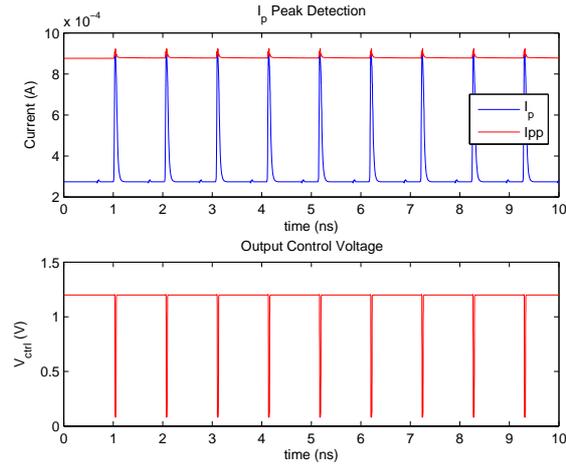


Figure 4.14: Peak Current Detection at 1GHz.

4.7 Circuit Performance of I_{DD} Sensors

The proposed I_{DD} based aging sensor was implemented using TSMC 65 nm technology to analyze its performance. Figure 4.14 depicts the transient waveform of the I_{DD} current and its afferent peaks. The experimental results have demonstrated that the proposed CPD can run at a frequency of 1 GHz, which stands for the applicability that how fast the proposed sensor can be utilized to perform a reliability assessment on the CUO. In practice, in order to assess the reliability status from a circuit path inside the CUO, multiple input patterns has to be injected into the circuit path to increase accuracy. Generally, multiple circuit blocks on a chip are under observation and some large circuit blocks might contain hundreds or thousands of circuit paths. Therefore, a high speed CPD circuit is very useful to the I_{DD} based aging sensor.

In order to assess the accuracy of the peak detector and current-to-time converter circuits, we use a two stage operational amplifier as test vehicle. The reliability analysis of NBTI and HCI induced aging is carried out by using Cadence RelXpert and Virtuoso Spectre simulators. Figure 4.15 presents current-to-time converting results and the error evaluation of the peak detector circuit.

The left axis represents the variation of the time delay T as a function of the control current I_{peak} for a load capacitance $C_1 = 1$ pF. For the purpose of illustration, we use a control current in the range 100 μ A to 1 mA, which

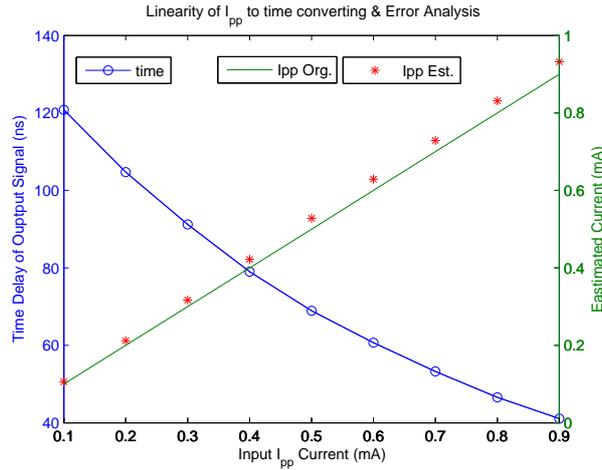


Figure 4.15: Linearity of Peak I_{DD} to Time Converting (left axis) and Error Analysis of Peak Detection (right axis).

results in a delay range of 120 ns to 40 ns. As one can observe in the figure, the converted time delay exhibits a fairly good linear relationship with the input I_{pp} current. The right axis represents the measured peak value of I_{DD} (" I_{pp} Est") compared with the ideal peak value (" I_{pp} Est"). The difference between the measure peak value and its corresponding ideal I_{pp} value represents the error introduced in the I_{pp} measurement process. From the experimental results, it is safe to conclude that a good accuracy on the I_{pp} current measurement can be achieved by the proposed I_{DD} based aging sensor.

In order to evaluate the validity of utilizing the peak value of the I_{DD} current as circuit aging monitor, we conducted accelerated testing simulation on the following ISCAS85 benchmark circuits: c499, which is a 32-bit single error correcting circuit comprising 202 gates and c880, which is an 8-bit ALU, comprising 383 gates.

The benchmark circuits are synthesized using the standard cells from the 45 nm NangateOpenCellLibrary [4] technology library. The reliability analysis is carried on by using Cadence RelXpert and Virtuoso Spectre simulators. As concerns the simulation environment, we employed several input aggression profiles consisting of different input patterns for each benchmark circuit. As environment parameters, we used a temperature of 27 °C, and a power supply $V_{DD} = 1.0$ V. We exposed the benchmark circuits to NBTI/PBTI and HCI wearout stress and adopted an end-of-life target of 10 years. For each benchmark circuit, we determined its critical path. Then we measured the degrada-

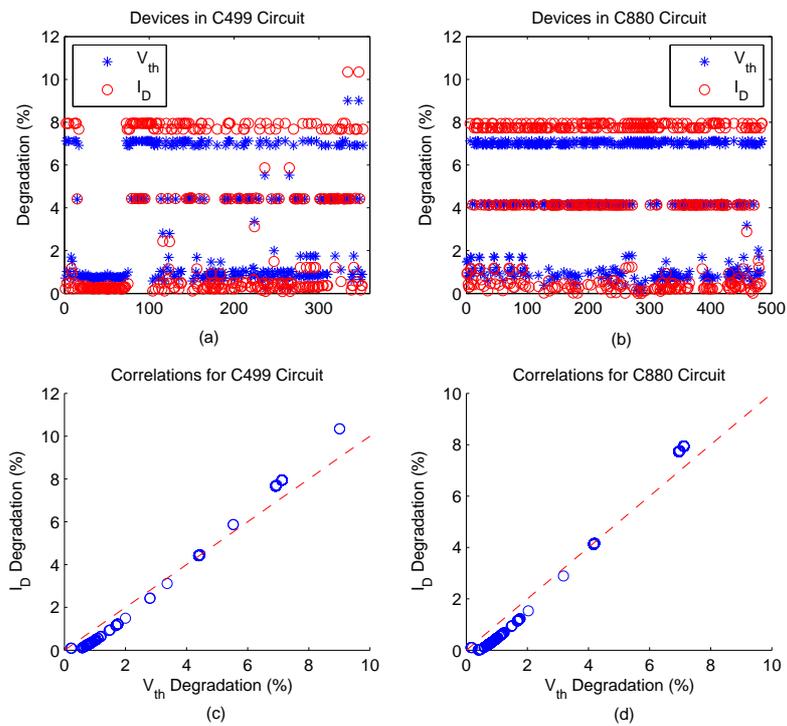


Figure 4.16: The percentage degradations of V_{th} and I_D for all devices in the c499 and c880 circuits - (a) and (b); and the correlations between the percentage degradations of V_{th} and I_D - (c) and (d).

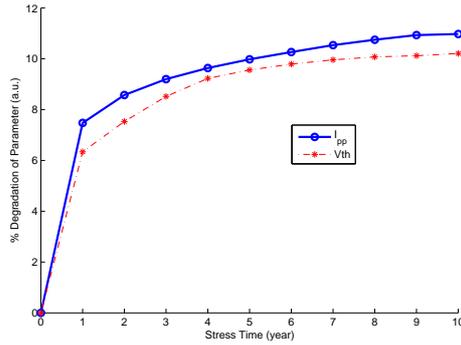


Figure 4.17: The Time Evolution of the V_{th} and I_D Degradation for 10-year Simulation.

tion of the V_{th} and the drain current I_D in percentage for every transistor on the critical path.

The V_{th} and I_D percentage degradation for all devices in the c499 and c880 circuits is graphically captured in Figure 4.16 (a) and (b). The x-axis is the device index and the y-axis parameter percentage degradation. It can be observed that for both considered circuits, for those devices which are less degraded (i.e., the percentage of degradation is small), the I_D degradation is smaller than the V_{th} degradation. As the degradation percentage becomes larger, the I_D degradation value increases faster than the V_{th} value and eventually, towards the conventional EOL (i.e., 10% degradation of the circuit critical parameters), it becomes larger than the V_{th} value. The improved sensitivity can be attributed to the dependence of I_D on multiple aging critical parameters, such as the threshold voltage V_{th} and the carrier mobility μ . This means that the I_D degradation could be a better indicator than the V_{th} degradation is, if we are concerned with the reliability status of mature which are approaching the final stage of their "career".

Figure 4.16 (c) and (d) depict the correlations between the V_{th} and I_D percentage degradations. The figures clearly indicate that the percentage degradation of I_D and V_{th} are strongly correlated, which means that both of them indicate the same aging trend for all the devices.

Figure 4.17 presents the time evolution curve for the V_{th} and I_{DD} degradation for 10-year simulation. As indicated in the figure, the I_{DD} curve has the same aging trend with the V_{th} curve, which confirms the validity of using I_{DD} as aging quantifier during the entire circuit lifetime.

4.8 Conclusion

In this chapter, we propose a novel Dynamic Reliability Management (DRM) framework which relies on quantitative degradation measurement. In order to implement such a DRM system, we propose two types of aging sensor to extract aging information from devices/circuits, i.e., V_{th} -based aging sensors for NBTI and HCI degradation measurement, and I_{DD} -based aging sensor for amalgamated degradation measurement, respectively. The V_{th} -based aging sensors are designed to be PVT-variation resistant. Our simulation results, using TSMC 65nm technology library, indicate that the proposed V_{th} sensors have good tolerance to dynamic variation with a sensitivity $0.29mV/^{\circ}C$ to temperature and $0.24mV/mV$ to voltage, respectively. The measurement deviation is maximum 7% for extreme conditions and negligible around the standard V_{DD} value. As a comparison, the V_{DD} -variation sensitivity is about $0.23 mV/mV$ for the NBTI sensor and $0.25 mV/mV$ for the HCI sensor in [51], with a deviation ranges from 8% to 18% and from 22% to 35% for the NBTI and HCI sensors, respectively, which demonstrates that our designs outperform state of the art equivalent counterparts.

The I_{DD} -based aging sensor is implemented in TSMC 65 nm technology library. The validity of using the I_{DD} value to monitor the circuit aging, is analyzed and experimentally verified by means of simulation for a set of ISCAS85 benchmark circuits. When compared to conventional V_{th} aging monitor, I_{DD} exhibits a better sensitivity as the circuit under aging assessment approaches its end-of-life, confirming the validity and practical relevance of the proposal to utilize it in aging monitoring frameworks.

The proposed aging sensors (V_{th} based or I_{DD} based) are mainly designed to assess reliability status from combinational logic blocks. In the next chapter, we will present self-adaptive compensation techniques for SRAM arrays to combat the aging-induced performance and stability degradations.

Note. The contents of this chapter is based on the the following papers:

Y. Wang, M. Enachescu, S. D. Cotofana, L. Fang, Variation Tolerant On-Chip Degradation Sensors for Dynamic Reliability Management Systems, Microelectronics Reliability, 2012(52), pp. 1787– 1791.

N. Cucu Laurenciu, Y. Wang, S. D. Cotofana, A Direct Measurement Scheme of Amalgamated Aging Effects with Novel On-Chip Sensor, to appear in the proceedings of IPF/IEEE International Conference on Very Large Scale Integration (VLSI-SoC), 2013.

5

Self-Adaptive Compensation Techniques for Independent Gate FinFET SRAM

As planar MOSFET is approaching its physical scaling limits, FinFET becomes one of the most promising alternative structure to keep on the industry scaling-down trend for future technology generations of 22 *nm* and beyond. In this chapter, we investigate the influence of NBTI degradation induced variation and random process variations on the stability of the FinFET based 6T-SRAM cell. The contributions of transistor threshold voltage variation ΔV_{th} on the stability of the SRAM cell and the corresponding compensating bias schemes are thoroughly examined by means of SPICE simulations. A mitigation method for memory stability management under spatial and temporal variations is demonstrated by taking advantage of the independent-gate FinFET device structure in order to perform threshold voltage adjustment. The proposed technique allows for a practical compensation strategy able to preserve the SRAM cell stability while balancing performance and leakage power consumption. We demonstrate that the standby leakage current I_{DDQ} value can be utilized to assess the consequences of parameter variations and NBTI on the circuit performance and propose a model that captures this. We evaluate the impact of our proposal on the SRAM cell stability by means of SPICE simulations for 20 *nm* FinFET devices. Simulation results indicate that the proposed technique can effectively maintain stability of an SRAM array within the desired range during its operational life under both spatial and temporal variations, hence improve the system performance and reliability. Our method allows for maintaining the Static Noise Margin (SNM) degradation of SRAM cells under a certain range, e.g., 2% of fresh device after 1 year operation, which is about 55.56% improvement when compared with the 4.5% degradation corresponding to the uncompensated case.

5.1 Introduction

As technology scaling continues, the Integrated Circuits (IC) feature size has been driven into the physical limitation edge of conventional MOSFET devices [3]. In order to keep technology scaling down further, two major issues have to be properly addressed: (i) the excessive leakage power, and (ii) the device/circuit reliability. Given that those issues have to be dealt with in the context of uncontrollable statistical process variations, the continuation of technology scaling seems to be more difficult than ever.

In order to keep on the industry scaling-down trend, novel devices and structures were proposed as potential candidates to replace the conventional planar MOSFET device [3, 92]. Among these devices, FinFET seems to be the most promising alternative structure for future technology generations of 22 nm and beyond, owing to its fabrication process simplicity and good electrical characteristics [92]. Due to its structure consisting of single/multiple vertical fin(s) FinFET has better electrostatics on Short-Channel Effect (SCE), thus less static leakage current and power consumption. However, FinFET is also sensitive to temporal degradations mechanisms, e.g., Negative Bias Temperature Instability (NBTI), Hot Carrier Injection (HCI). In other words, on top of the spatial uncertainty caused by process variations, also for FinFET various temporal degradations hold back the feature size of technology from further scaling.

In view of the previous argument, in this chapter, we address the following aspects:

- Modeling of SRAM cell stability under spatial and temporal V_{th} variations induced by process variation and NBTI stress;
- Dynamic characterization of NBTI induced V_{th} degradation by monitoring the standby leakage current I_{DDQ} ;
- Mitigation methods able to compensate the effects of process variations and NBTI by taking advantage of the FinFET's special device structure.

We evaluated our proposals by SPICE simulations for 20 nm FinFET devices, and the results indicate that the proposed technique can effectively maintain stability of SRAM array within the desired range during its operational life under both spatial and temporal variations. For example, our method allows for maintaining the SNM degradation of FinFET SRAM cells under 2% of the fresh device SNM after 1 year operation, which is about 55.56% improvement when compared with the 4.5% SNM degradation corresponding to the

uncompensated case.

The rest of this chapter is organized as follows: In Section 5.2 we introduce the temporal V_{th} degradation induced by NBTI under dynamic stress; In Section 5.3 we analyze the impact of V_{th} variations caused by process variations and NBTI degradation on the stability of SRAM cells; In Section 5.4 we propose a V_{th} compensation technique using independent double-gate FinFET to improve SRAM's stability and the proposed scheme is evaluated in Section 5.5 by means of SPICE simulations; In Section 5.6 we concludes this chapter.

5.2 Temporal and Spatial V_{th} Variations

The Negative Bias Temperature Instability (NBTI) phenomenon and its consequences have been extensively studied [13, 16, 111] and there is clear indication that NBTI caused degradation is becoming a major reliability concern for nanoscale CMOS technology. NBTI is prominent in PMOS devices and it causes a threshold voltage (V_{th}) shift, which results in poor drive current and in shorter device and by implication circuit lifetime.

NBTI occurs along the entire transistor channel at elevated temperature when negative gate-to-source voltage is applied. Holes from the inversion layer can tunnel into the gate oxide, break the Si-H bond leaving behind an interface trap, which results in a positive V_{th} shift. Traditionally, the interface trap generation is modeled within the Reaction-Diffusion (R-D) framework [13], which gives a power-law time evolution of V_{th} degradation. A long-term V_{th} shift under dynamic stress is given by [16]:

$$\Delta V_{th} = At^n = \left(\frac{\sqrt{K_v^2 \alpha T_{clk}}}{1 - \beta_t^{1/2n}} \right)^{2n}, \quad (5.1)$$

where $n = 1/6$ is the power-law time constant, T_{clk} is the clock period, α ($0.1 \leq \alpha \leq 0.9$) is the NBTI stress probability, i.e., the NBTI duty-cycle. β_t is a coefficient reflecting the NBTI recovery effect and is computed as follows:

$$\beta_t = 1 - \frac{2\xi_1 t_e + 2\sqrt{\xi_2 C(1 - \alpha) T_{clk}}}{2t_{ox} + \sqrt{Ct}}, \quad (5.2)$$

where ξ_1 , ξ_2 , and C are fitting constants, t_{ox} is the oxide thickness, t_e is the effective diffusion distance of hydrogen species. K_v stands for technology, supply voltage, and operating temperature dependence associated with NBTI

degradation. At its turn K_v is computed as [16]:

$$K_v = \left(\frac{qt_{ox}}{\epsilon_{ox}} \right)^3 K_1^2 C_{ox} (V_{gs} - V_{th}) \sqrt{C} \exp\left(\frac{2E_{ox}}{E_0}\right), \quad (5.3)$$

where q , k , ϵ_{ox} , K_1 , C_{ox} , V_{gs} , E_{ox} , and E_0 are the elementary charge, Boltzmann's constant, oxide permittivity, constant factor, oxide capacitance per area, gate-source voltage, vertical oxide field, and fitting constant, respectively.

Apart of NBTI, the V_{th} is also influenced by Process Variations (PV), which induce spatial uncertainties on the device performance relevant parameters. Process variations become particularly important in smaller technology nodes ($< 65 \text{ nm}$) as when the feature size is scaling down the variation consequences are becoming more significant and have larger impact on device size and performance. Process variations are typically divided into two components: (i) inter-die (global), which accounts for chip- or wafer-level variations and (ii) intra-die (local), which accounts for variations between different devices in the same die.

For simplicity, we assume that the PV-induced ΔV_{th} is static and the time evolution of V_{th} can be expressed as follows:

$$V_{th} = V_{th0} + \Delta V_{th}^g + \Delta V_{th}^l + \Delta V_{th}^{nbt_i}(t), \quad (5.4)$$

where V_{th0} is the nominal V_{th} value, $\Delta V_{th}^{nbt_i}$ is the NBTI-induced V_{th} degradation, and ΔV_{th}^g and ΔV_{th}^l are V_{th} alterations due to global and local variations, respectively.

As device dimension scale down into nanometer region, Random Dopant Fluctuation (RDF) becomes one of the major variations affecting the performance of device. RDF directly affects the threshold voltage of a MOSFET, since V_{th} depends on the charge of the ionized dopants in the depletion region [9]. According to [99], the variance of V_{th} mismatch caused by RDF follows a Gaussian distribution, and its standard deviation can be modeled as:

$$\sigma_{\Delta V_{th}, RDF} = \frac{t_{ox}}{\epsilon_{ox}} \cdot \frac{\sqrt[4]{2q^3 \epsilon_{si} N_a \phi_B}}{\sqrt{3} W_{eff} L_{eff}}, \quad (5.5)$$

where N_a is the channel doping concentration, ϵ_{si} is the *Si* permittivity, ϕ_B is the difference between the Fermi Level and the intrinsic level, and $W_{eff} L_{eff}$ denotes the transistor's active area.

The PV-induced V_{th} variations define the statistical reliability profile of the circuit, SRAM cells in our case, at time 0, i.e., when devices are fresh. As

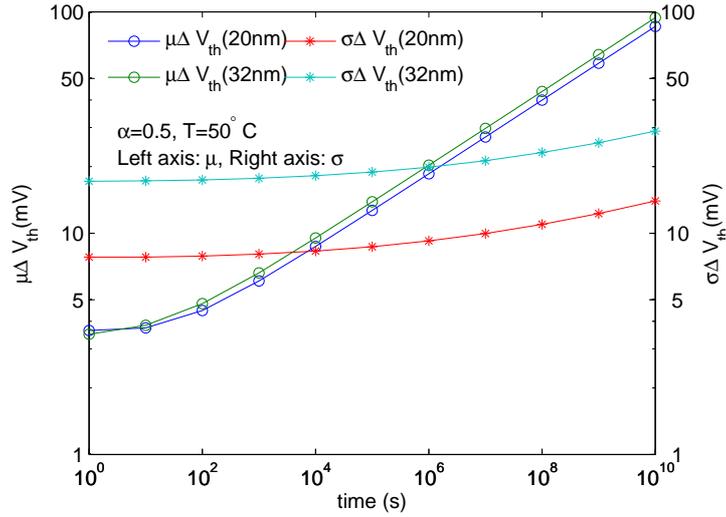


Figure 5.1: NBTI-Induced V_{th} Degradation Mean Value and Standard Deviation for PTM 32 nm and 20 nm/ FinFET Library [5].

known, NBTI induces a temporal V_{th} degradation and, for an individual device, V_{th} evolution in time is governed by Eq. (5.1). However, due to the RDF effect, the Si-H bonds dissociation and re-passivation processes experience stochastic fluctuations. Thus, by taking the PV influence into consideration, the long term V_{th} degradation induced by NBTI can be expressed as [113]:

$$\Delta V_{th}(t) = A(1 - S_v(\Delta V_{th}^g + \Delta V_{th}^l))t^n, \quad (5.6)$$

where S_v is a threshold voltage sensitivity coefficient. As a result, the variation of V_{th} temporal shift due to NBTI, given its mean $\mu(\Delta V_{th}(t))$, can be expressed as [90]:

$$\sigma_{\Delta V_{th}, NBTI} = \sqrt{\frac{2q}{\epsilon_{si}} \cdot \frac{t_{ox} \mu(\Delta V_{th}(t))}{W_{eff} L_{eff}}} \propto t^{1/12}. \quad (5.7)$$

Finally, the variation of V_{th} when considering both mismatch and NBTI effects can be calculated as

$$\sigma_{\Delta V_{th}} = \sqrt{\sigma_{\Delta V_{th}, RDF}^2 + \sigma_{\Delta V_{th}, NBTI}^2}. \quad (5.8)$$

Figure 5.1 represents the calculated mean value and standard deviation of the NBTI-induced ΔV_{th} degradation according to the discussion above. After a

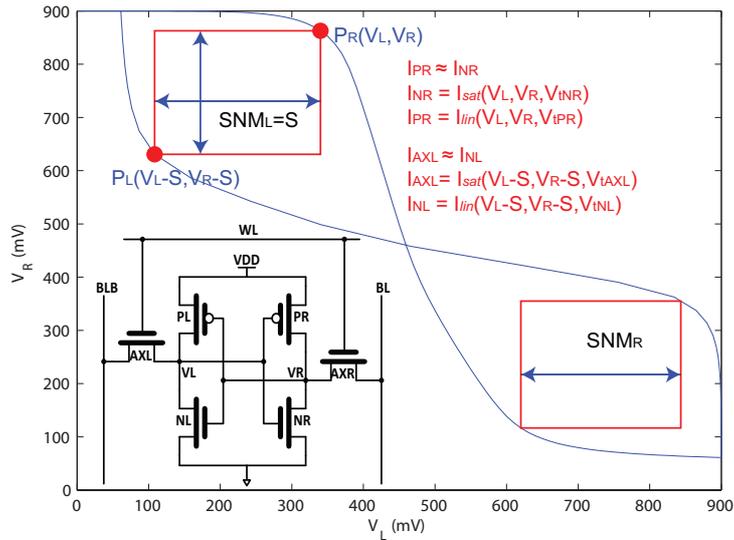


Figure 5.2: 6T SRAM Cell Schematic and Butterfly Curve (PTM 32 nm Technology, $V_{DD} = 0.9$ V).

short stressing period ($t > 10^2$ s), the power-law dependence can be clearly observed for both PTM 32 nm planar and 20 nm FinFET technologies [5].

The NBTI failure mechanism not only shifts the V_{th} parameter, but also increases the spread of its value. Hence, the influence of device parameter variation on the circuit performance should be taken care dynamically.

5.3 ΔV_{th} Impact on SRAM Cell Stability

As a major part of modern processors, SRAM drives the technology scaling direction in industry. Hence, ensuring the SRAM cell reliability is critical for technology scaling, as due to its small area and power consumption it is more sensitive to temporal degradation and spatial variations than other components. In view of this in the following we study the influence of threshold voltage value and variation on the Static Noise Margin (SNM), which is a metric reflecting the memory cell stability.

5.3.1 SNM vs. ΔV_{th}

A conventional 6T SRAM cell and its stability diagram are presented in Figure 5.2. The most critical SNM is READ SNM, since READ operation has a more severe condition than HOLD and WRITE operations. The SNM (for READ, hereinafter if without special specification) can be derived by solving the Kirchhoff's Current Law (KCL) at the cell storage nodes V_L and V_R for read operation, respectively:

$$\begin{aligned} I_{NR} &= I_{PR} + I_{AXR}, \\ I_{NL} &= I_{PL} + I_{AXL}. \end{aligned} \quad (5.9)$$

For the simplicity of discussion, the drain current can be estimated by an alpha-power law current model [93]. If we assume that $V_L \approx V_{dd}$ and $V_R \gg V_{tPL}$ for the left side SNM, then I_{AXR} and I_{PL} are negligible. As depicted in Figure 5.2, in the PR and PL neighbourhood solving the KCL equations for V_R yields:

$$\begin{aligned} V_{dd} - \frac{\beta_{NR}}{\beta_{PR}} \frac{(V_L - V_{tNR})^\alpha}{(V_{dd} - V_L - V_{tPR})} = \\ V_{tNL} + S + \frac{\beta_{AXL}(V_{dd} - V_L + S - V_{tAXL})^\alpha}{\beta_{NL}(V_L - S)}, \end{aligned} \quad (5.10)$$

where V_L is the node voltage at left SRAM node when $S = SNM_L$ reaches maximum, α is the constant from alpha-power law current model [93], and $\beta_x = \mu_{eff} C_{ox} W_x/L$ are the coefficients in the current equation and V_{tx} are the threshold voltages, where $x \in \{AXL, NL, PR, NR\}$, as depicted in Figure 5.2. A similar relationship can be derived for the SNM of right node.

From Eq. (5.10) we can deduce that SNM_L is determined by four transistors, namely PR, NR, AXL, and NL. Hence, the SNM_L fluctuation is a function of V_{th} variations of these four transistors, i.e.,:

$$\Delta SNM_L = \sum_x \frac{\partial SNM_L}{\partial V_{tx}} \cdot \Delta V_{tx}, \quad (5.11)$$

where $x \in \{PR, NR, AXL, NL\}$.

Applying partial derivative on both sides of Eq. (5.10), we can solve the dependence of SNM_L on the V_{th} variations of the corresponding transistors, which can be expressed as:

$$\frac{\partial SNM_L}{\partial V_{ti}} = k_i(V_{dd}, \beta, V_{tj}). \quad (5.12)$$

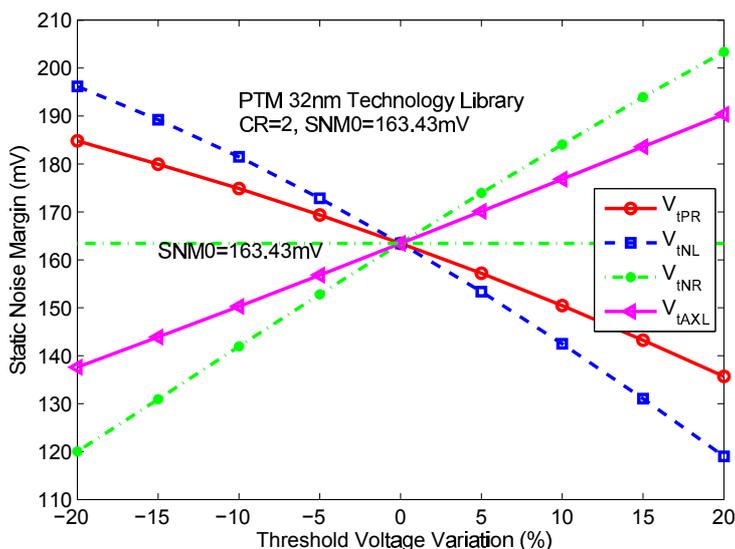


Figure 5.3: SNM_L vs. ΔV_{th} Variations for the 6T SRAM Cell Transistors (PTM 32 nm Planar Devices, $CR=\beta_{NL}/\beta_{AXL}=2$).

where $j \in \{PR, NR, AXL, NL\} \cap j \neq i$. We can assume that the V_{tj} are constant for a specific technology, then k_i is constant. Hence, SNM_L has a linear dependence on the V_{th} variations. We verified this relationship by means of SPICE simulation, the results are presented in Figure 5.3 for a PTM 32nm planar-device technology library.

5.3.2 SNM vs. ΔV_{th} Modulation

In order to control the variation-induced stability fluctuations in the 6T-SRAM cell, Adaptive Body Bias (ABB) can be dynamically applied to modulate the V_{th} value of the corresponding transistors, i.e., transistors PR, NR, AXL, and NL for SNM_L . Notice that SNM_L has positive depending-coefficients on ΔV_{th} in transistor AXL and NR, and has negative depending-coefficients on ΔV_{th} in transistor PR and NL. Hence, to compensate the V_{th} -variation induced SNM fluctuations, Forward Body Bias (FBB) is required for AXL and NR; and Reverse Body Bias (RBB) is required for PR and NL.

Furthermore, to increase SNM_L , high V_{tNL} and low V_{tNR} are of interest. Symmetrically, to increase SNM_R , low V_{tNL} and high V_{tNR} should be targeted. As a result, V_{tNL} and V_{tNR} are not suitable to be utilized as compensating parameters in a symmetric design. This conclusion is illustrated in Figure 5.4. As

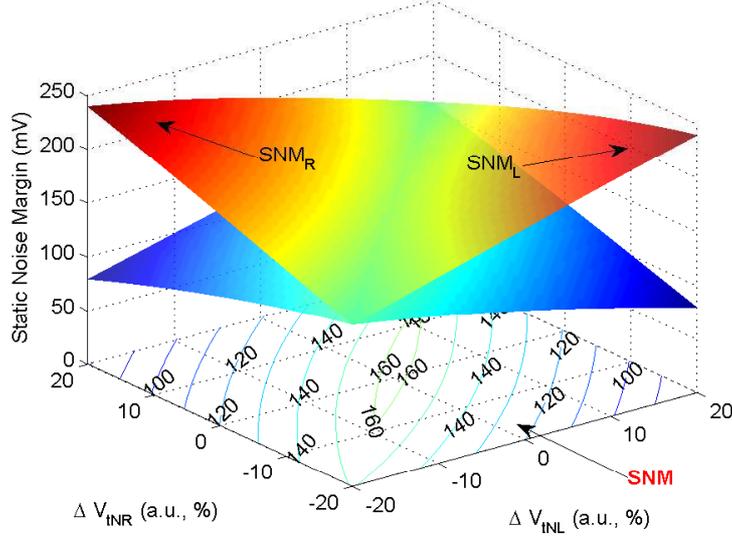


Figure 5.4: SNM vs. ΔV_{th} Variations (in the NL, NR Transistors). The contour lines on the bottom plane represent the overall $SNM = \min(SNM_L, SNM_R)$ of the cell.

we can observe in the figure, the optimized overall SNM of a cell ($SNM = \min(SNM_L, SNM_R)$) is located at the point where $\Delta V_{tNL} = \Delta V_{tNR} \approx 0$, which means zero bias is always preferred.

With novel device structures specific to emerging technologies, we have opportunities to compensate/suppress the variation-induced stability fluctuations more efficiently. In the next section, we introduce a variation suppression and mitigation technique for SRAM array using double-gate devices.

5.4 IG-FinFET SRAM Stability Mitigation

In an independent-gate configuration of FinFET (IG-FinFET) [92], a separated back-gate can be used to control the threshold voltage of the device. The "front-coupling" dependence of the threshold voltage on the back-gate voltage can be expressed as [50]:

$$\gamma_b = \frac{\partial V_{th}}{\partial V_{bg}} = -\frac{C_{si} C_{ox2}}{C_{ox1}(C_{si} + C_{ox2})}, \quad (5.13)$$

where $C_{ox1/2} = \epsilon_{ox}/t_{ox1/2}$ and $C_{si} = \epsilon_{si}/w_{si}$ are the front- and back-gate oxide capacitance and body capacitance, respectively. The negative sign in

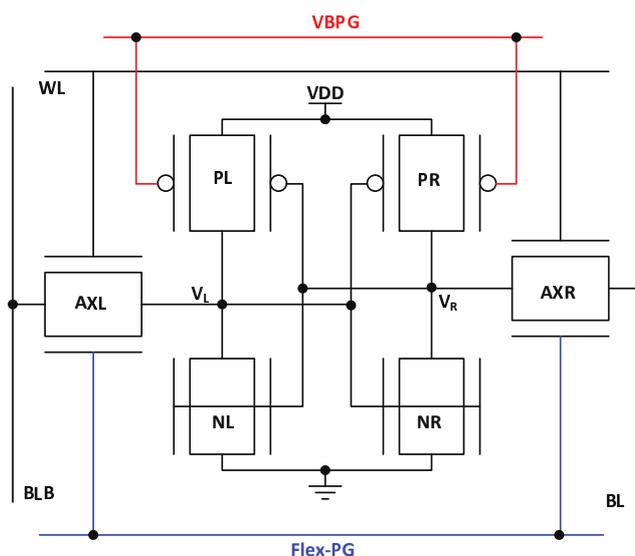


Figure 5.5: IG-FinFET 6T SRAM with V_{th} Compensation/Adjustment for PMOS and Pass Gates: $VBPG$ compensates the PMOS NBTI-induced V_{th} degradation, and $tFlex-PG$ bias adjusts the V_{th} of the pass gates to improve the SRAM cell stability.

Eq. (5.13) indicates that the direction of threshold voltage change is opposite to that of the back-gate voltage change.

5.4.1 IG-FinFET Based V_{th} Compensation Scheme

Figure 5.5 presents the configuration of a IG-FinFET based 6T SRAM cell with V_{th} compensation for PMOS (through the extra bias $VBPG$) and V_{th} adjustment for the pass gates AXL and AXR (through the extra bias $Flex-PG$) to improve its stability. The $Flex-PG$ technique was proposed in [81] to improve the read and write stability by adjusting the V_{th} of the pass gates. A high V_{th} is desirable for READ stability and a low V_{th} is preferred to improve WRITE stability. This relationship is illustrated in Figure 5.6. During the read operation, the access transistors are forward biased to increase the READ SNM; while during the write operation, they are reverse biased to increase the WRITE SNM. This relationship between stability modulation and the value of the applied bias is straight forward, which simplifies the strategy for performance management. In our proposal, the $Flex-PG$ range is fixed to compensate the PV-induced SNM fluctuations. The exact $Flex-PG$ value is determined by a global PV-sensor, and this value is common to all SRAM arrays in a die.

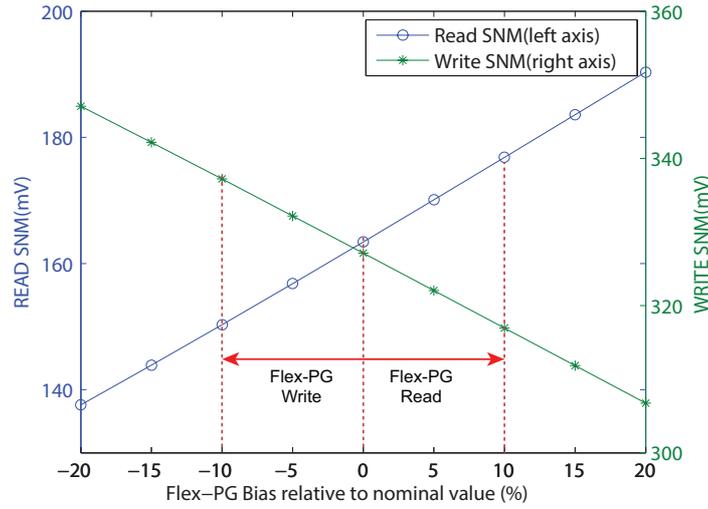


Figure 5.6: *Flex-PG* vs. SRAM Read/Write Stability.

The time-dependent stability fluctuation induced by NBTI is compensated by the *VBPG* bias. Figure 5.7 depicts the NBTI-induced SNM degradation versus the ΔV_{th} of the two PMOS devices in the SRAM cell. As one can observe in the figure, the amount of SNM degradation is sensitive to the signal probability α (i.e., the probability that the internal left/right node stores 0). The least degradation path (along which the SNM contour line decrease is the slowest) locates at $\alpha = 0.5$, which means that the left and the right node have an equal probability to store 0. A cell-flip technique was proposed in [58] to balance the signal probabilities between the two SRAM cell nodes. This proposal is efficient to slow down the NBTI induced degradation, however, it introduces a large area overhead and performance penalty for implementation. Instead of node-balancing, we employ a common bias for both nodes to compensate the performance degradation. As illustrated in Figure 5.7, with the *VBPG* bias applied, the degradation can be compensated or even eliminated for the extreme unbalanced cases (i.e., $\alpha = 0.1$ or 0.9).

5.4.2 V_{th} Compensation Using Supply Leakage Current Monitoring

In [46] the authors have analyzed and proposed to use the standby leakage current I_{DDQ} to monitor and characterize the NBTI induced temporal perfor-

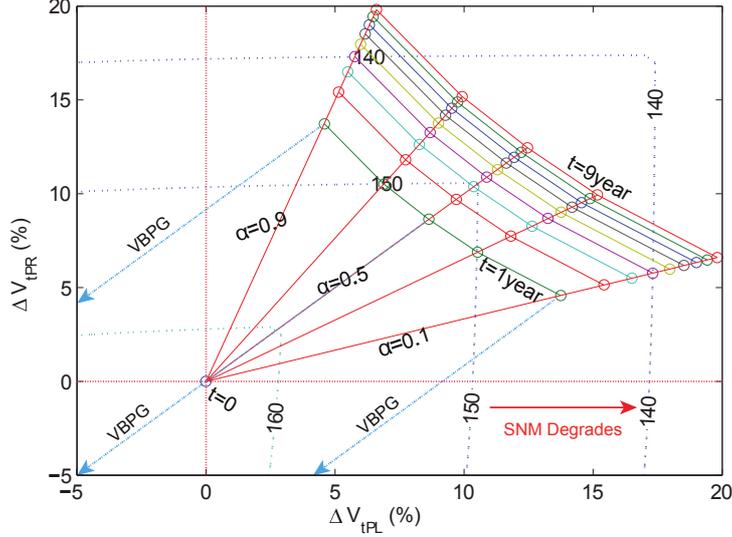


Figure 5.7: NBTI-induced SRAM cell SNM degradation presented in the $\Delta V_{tPL} \times \Delta V_{tPR}$ plane. The dashed lines with number labels are contour lines for overall SNM of the two nodes. The color lines with cycles are the degraded SNMs after 1 to 9 year(s), and the solid straight red lines represent the SNM time evolution for a given signal probability α at the left node of the SRAM cell.

mance degradation. As suggested by their work, a current sensor monitoring the I_{DDQ} for the entire SRAM array is a good indicator of the NBTI induced degradation. In this section we further extend this idea in order to introduce an NBTI mitigation technique.

The I_{DDQ} of a circuit is defined as the total leakage current in standby mode, which for an SRAM array with N cells can be expressed as follows:

$$I_{DDQ} = \sum_{i=1}^N I_{DDQi} = \sum_{i=1}^N I_0 \exp\left(-\frac{V_{ti}}{m v_T}\right), \quad (5.14)$$

where $I_0 = \beta(m-1)(1 - \exp(-V_{ds}/v_T))$, m is the body effect coefficient, and v_T is the thermal voltage (kT/q). Under the assumption that the ΔV_{th} due to RDF and NBTI follows a Gaussian's distribution. The leakage current

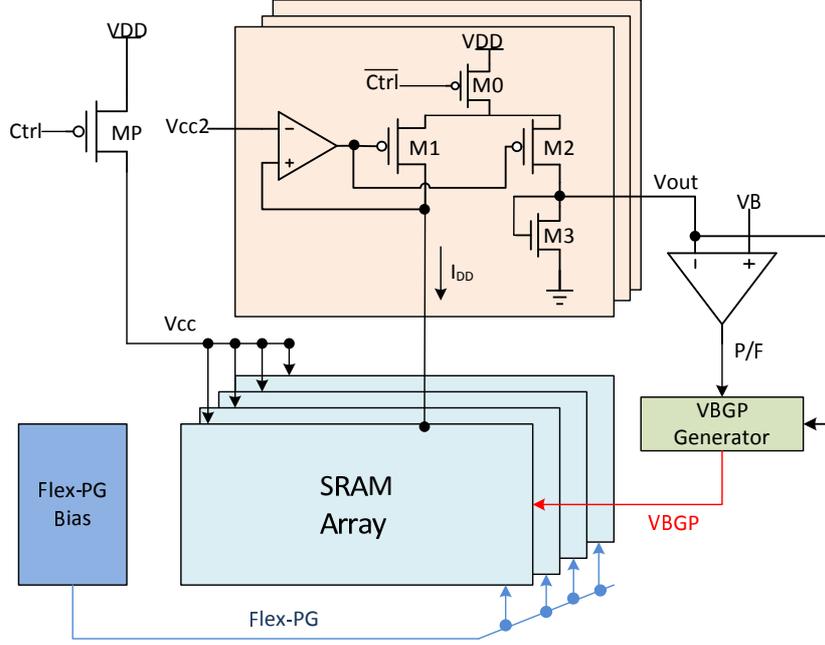


Figure 5.8: NBTI Mitigation Using I_{DDQ} Monitor.

value in SRAM cells follows a Log-Normal distribution [76], which gives:

$$\mu(I_{DDQi}) = I_0 \exp\left(\frac{-\mu + \sigma^2/2}{mV_T}\right), \quad (5.15)$$

$$\sigma^2(I_{DDQi}) = I_0 \left(\exp\left(\frac{\sigma^2}{mV_T}\right) - 1\right) \exp\left(\frac{-2\mu + \sigma^2}{mV_T}\right), \quad (5.16)$$

where μ and σ are the Mean and Standard Deviation of the V_{th} value, respectively.

According to the Central Limit Theorem, the summation of independent random variables (e.g., I_{DDQi}) can be assumed to follow a Normal Distribution, thus if N is a large number, the total standby leakage current and its standard deviation can be expressed as:

$$\begin{aligned} \mu(I_{DDQ}) &= \sum_{i=1}^N \mu(I_{DDQi}), \\ \sigma^2(I_{DDQ}) &= \sum_{i=1}^N \sigma^2(I_{DDQi}). \end{aligned} \quad (5.17)$$

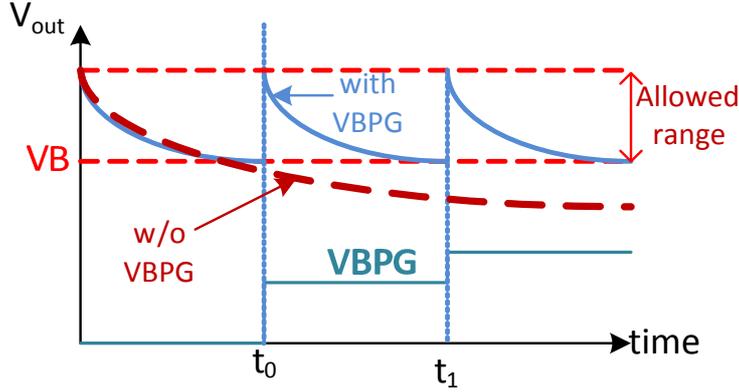


Figure 5.9: A Practical Compensation strategy for NBTI Induced SNM Degradation.

From Eq. (5.4), (5.6), (5.7), (5.8) and (5.17) we get:

$$\mu(I_{DDQ}) = I_{DDQ0} \cdot \exp\left(-\frac{K_n \Delta V_t(t)}{m v_T}\right), \quad (5.18)$$

$$\sigma(I_{DDQ}) = \sqrt{N} \sigma(I_{DDQi}). \quad (5.19)$$

where $K_n = (1 + q t_{ox}) / \epsilon_{si} W L$ and

$$I_{DDQ0} = N I_0 \cdot \exp(-(V_{th0} + \delta) / m v_T),$$

$$\delta = \Delta V_t^g + \sigma \Delta V_{t,RDF}^2 / 2,$$

where V_{t0} is the nominal value of threshold voltage.

Eq. (5.18) suggests that the total standby leakage current I_{DDQ} decreases exponentially with time due to NBTI effect, hence, it is a good indicator to monitor the NBTI-induced degradation. The specific design of the proposed NBTI mitigation scheme using I_{DDQ} monitoring is depicted in Figure 5.8. In the proposed scheme, a I_{DDQ} current sensor formed by the transistors $M1 \sim M3$ is attached to each and every SRAM array. A signal called $Ctrl$ is used to toggle the working mode of the sensor: during normal operation, the current sensor is bypassed through the transistor MP ; while in measure mode, MP is cut off and the power supply current of the SRAM array under monitoring is forced to flow through $M1$. The current is then mirrored into $M2$ and is converted into a voltage signal by $M3$, generating an output signal V_{out} . The output signal is subsequently compared with a reference voltage VB to evaluate the severity of the NBTI induced degradation. The comparison result is utilized to activate the $VBGP$ generator, which produces a proper bias and sends it to the back-gate

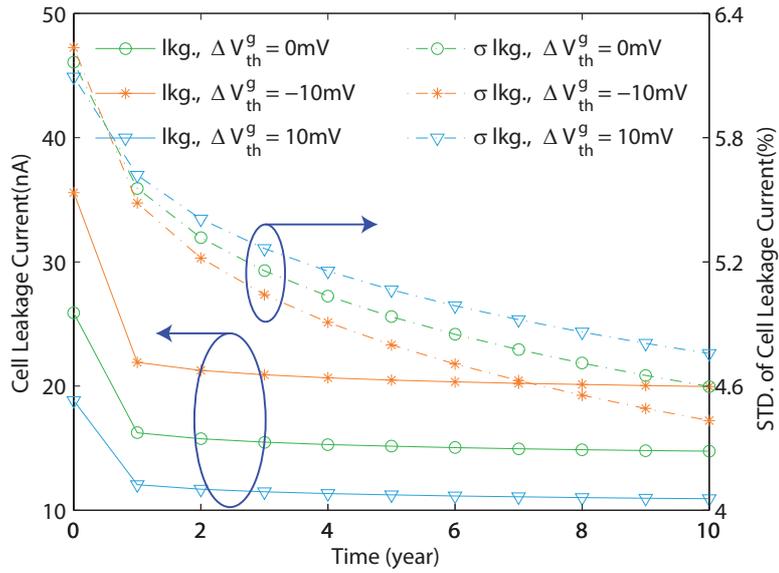
of the PMOS devices in the SRAM array. The I_{DDQ} current is measured when the SRAM is in standby, i.e., the bit lines BL and BLB are precharged and the word line WL is set to "0". Considering that the NBTI induced degradation is a relatively slow progress, the value of V_{th} compensation for the SRAM array needs to be calibrated only every now and then, thus the SRAM array can be almost all the time in normal operation mode.

The controlling bias VB in Figure 5.8 determines the actual compensation strategy according to the information provided by the I_{DDQ} sensor. Since the magnitude of the output signal V_{out} is designed to be proportional to the leakage current I_{DDQ} , the compensation strategy is a trade off between performance and leakage power. A simple but efficient compensation strategy in practice is to set a lower limit for the acceptable I_{DDQ} value. In this way, the performance degradation is set to be in an allowed range, which is determined by the VB value, to trade performance for the leakage reduction induced by NBTI. When the performance degradation exceeds the predefined range, the compensation scheme is activated to bring back the performance in the desired range. This strategy is demonstrated in Figure 5.9: when the readout of I_{DDQ} V_{out} reaches the pre-defined bottom line VB , a corresponding $VBPG$ is assigned to compensate the NBTI-induced degradation. This $VBPG$ is kept until V_{out} reaches VB again.

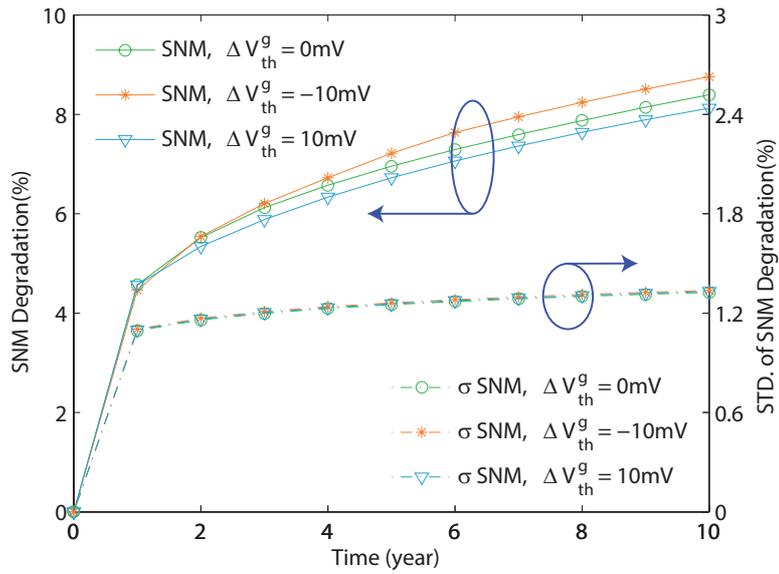
5.5 Simulation Results

In order to evaluate the efficiency of the proposed compensation technique, we run circuit simulations using the 20nm PTM library for FinFET [5]. In this library, the BSIM CMG [2] model is utilized for simulation. As up to our best knowledge no public SPICE model is available for IG-FinFET devices, we use the V_{th} deviation parameter to simulate the V_{th} modulation by back-gate bias. This simplification does not change the effectiveness of the proposed compensation technique, since back-gate modulation for V_{th} in IG-FinFET is verified by both simulations and experiments [66]. The PV variations are generated using a Gaussian distribution for RDF-induced V_{th} variation, as described by Eq. (5.5). The NBTI stress is set as 10-year operation at $50^\circ C$.

We first investigate the time evolution of the cell leakage current under NBTI stress. To demonstrate the influence of the global variation on the degradation, we considered three cases with $-10mV$, $0mV$, and $10mV$ global V_{th} variation, respectively. Figure 5.10(a) represents the NBTI-induced cell leakage current (I_{DDQ_i}) degradation versus time for the 10-year period. As one can

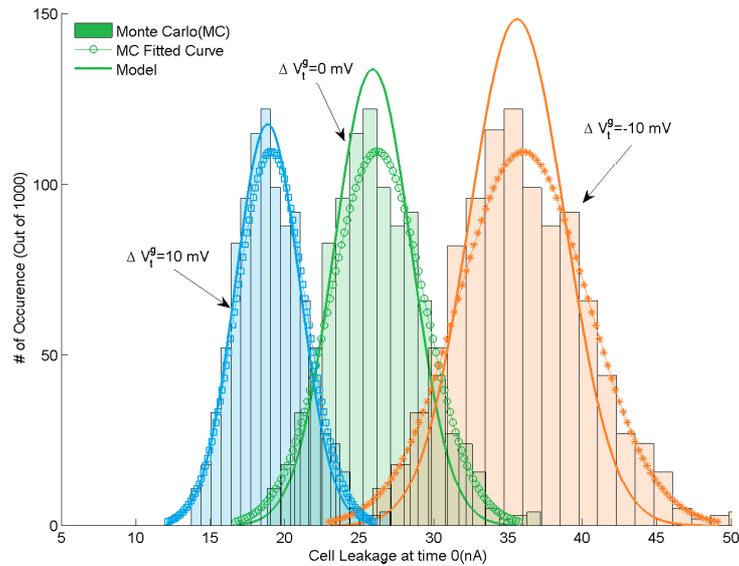


(a) Nominal I_{DDQ} ("Ikg." on the left axis) and its relative standard deviation (" $\sigma Ikg.$ " on the right axis) vs. Time

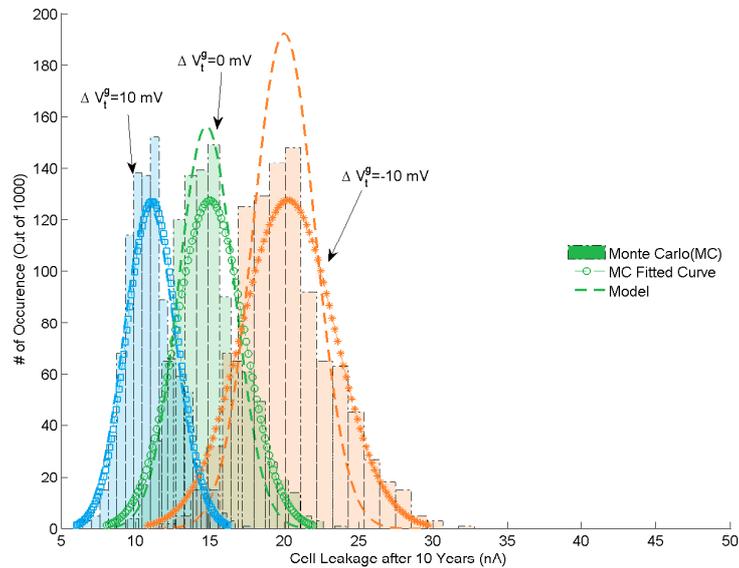


(b) Nominal SNM degradation relative to the fresh device SNM (left axis) and its relative standard deviation (right axis) vs. Time

Figure 5.10: 6T-SRAM cell standby leakage and SNM degradation (10-year operation time at $50^{\circ}C$) using 20nm FinFET Technology with $-10mV$, $0mV$, and $10mV$ global variations, respectively.



(a) Fresh Device



(b) After 10-year Operation

Figure 5.11: 6T-SRAM cell leakage distribution of fresh device and aged device(10 years operation at 50°C) using 20 nm FinFET technology with -10 mV , 0 mV , and 10 mV global variations, respectively.

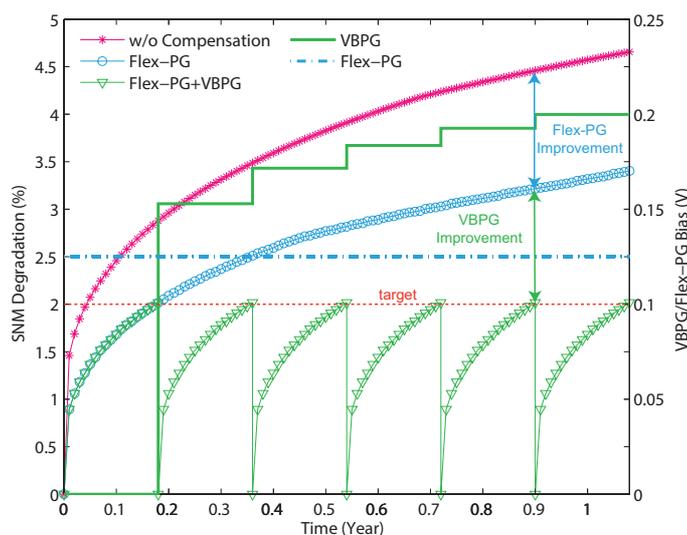


Figure 5.12: SNM improvement with the VBPG and Flex-PG compensating technique with symmetric double-gate 20nm FinFET technology: $t_{ox1}=t_{ox2}=1.4\text{nm}$, $w_{sj}=t_{fin}=15\text{nm}$.

observe in the figure, the I_{DDQ_i} current decreases very fast at the beginning-of-life, because the NBTI-induced ΔV_{th} follows a power-law of time and I_{DDQ_i} has an exponential dependence on ΔV_{th} . This I_{DDQ_i} feature makes it a good indicator to assess the NBTI-induced degradation at the beginning of the operational life, which is crucial to control IC's performance degradation for components like SRAM. Furthermore, the global variation has significant influence on the I_{DDQ_i} magnitude, but has little effect on the deviation, which means that using I_{DDQ_i} as degradation indicator is stable in the presence of global variations. Figure 5.10(b) presents the SNM degradation relative to the fresh device SNM and the deviation of SNM degradation relative to the SNM at corresponding time for the 10-year operation. The results suggest that ΔSNM follows a power-law rule, as described by Eq. (5.11). One can observe that the deviation of ΔSNM increases with time, which means that the performance uncertainty becomes larger at the end-of-life. As a result, the V_{th} compensation has to take this uncertainty into consideration for a heavily stressed component.

In order to evaluate the accuracy of the estimation model (i.e., Eq. (5.15) and Eq. (5.16)) on the cell leakage distribution, we run Monte-Carlo simulations for RDF and NBTI induced ΔV_{th} variations and the simulation results are pre-

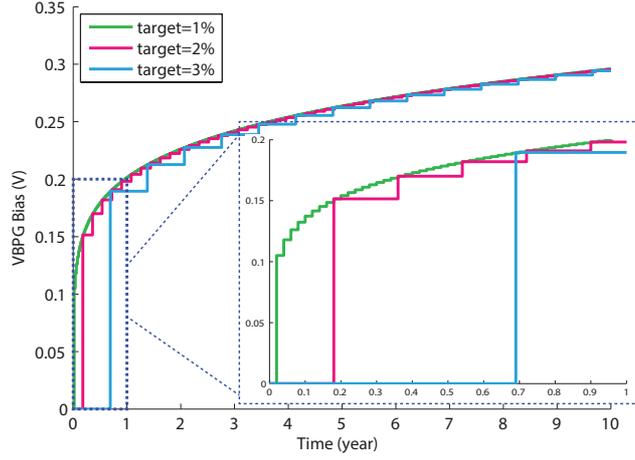


Figure 5.13: The Applied VBPG Bias with Different Targeted SNM Degradation Margins.

sented in Figure 5.11. The data in the figure clearly indicate that, the distributions of cell leakage are accurately captured by the proposed I_{DDQ_i} model for both fresh devices and aged devices after 10-year operation. We recall that the accurate estimation on the I_{DDQ_i} magnitude and spread range is critical to set the allowed performance degradation range for the ΔV_{th} compensation technique. Its underestimation leads to an increasing soft error rate in the SRAM array while its overestimation leads to a higher leakage power consumption.

Figure 5.12 presents the SNM improvement obtained by means of *VBPG* and *Flex-PG* compensation/mitigation techniques. The *VBPG* compensation trigger is set at 2% degradation of the fresh device SNM. A symmetric double-gate FinFET, with thickness of front- and back-gate oxide $t_{ox1}=t_{ox2}=1.4\text{nm}$, and fin thickness $w_{si}=15\text{nm}$ was utilized in the simulation. One can observe in the figure that, when compared with the degradation without any compensation technique, *Flex-PG* can reduce about 26.67% of the SNM degradation at the end of one year operation (from $\sim 4.5\%$ to 3.3% degradation versus fresh device SNM). *VBPG* compensation can maintain the SNM degradation after 1 year under 2% of the fresh device, which is about 55.56% improvement when compared with the uncompensated case. The magnitude of the applied forward bias *VBPG*, calculated by Eq. (5.13), is presented in the figure as well.

Figure 5.13 presents the required *VBPG* bias corresponding to different SNM degradation margin targets for a 10-year operation. As one can observe in the figure, different degradation targets set different resolutions for the compensa-

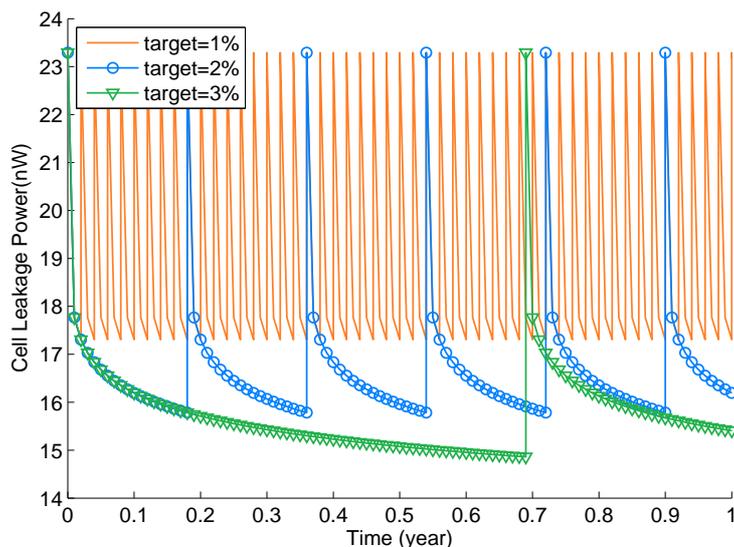


Figure 5.14: The Cell Leakage Power Consumption versus Different SNM Degradation Targets.

tion bias *VBPG*. As degradation increases, the *VBPG* bias for different targets saturate to the same value towards the end-of-life. Hence, a more intelligent compensation strategy is to set a fine-grain *VBPG*-calibration at the beginning-of-life to improve the cell stability, and a coarse-grain *VBPG*-calibration after certain operation time (e.g., 1 year) to increase the availability in normal operation mode.

From the cell stability point of view, high *VBPG* is preferred. However, high *VBPG* increases the standby leakage significantly. Figure 5.14 presents the average cell leakage power consumption corresponding to different SNM degradation targets. The average power consumption difference between the worst case (targeting 1% degradation) and the best case (targeting 3% degradation) could be as large as 20% (i.e., $19.45nW$ vs. $15.63nW$). Hence, the compensation strategy is a trade-off between cell stability and power consumption, and bias-calibration frequency and normal operation time as well.

For the simplicity of implementation, Eq. (5.18) implies a uniform NBTI duty-cycle " α " of the internal nodes (for the left and right node in a cell, $\alpha_L + \alpha_R \cong 1$) for cells in the SRAM block monitored by the same I_{DDQ} sensor. However, the cell duty-cycles are not uniform in practice. The real duty-cycle of a cell is dependent on the "0/1" value ratio (i.e., workload) stored in the SRAM cell. Other than previous works like [58] that introduces extra hardware

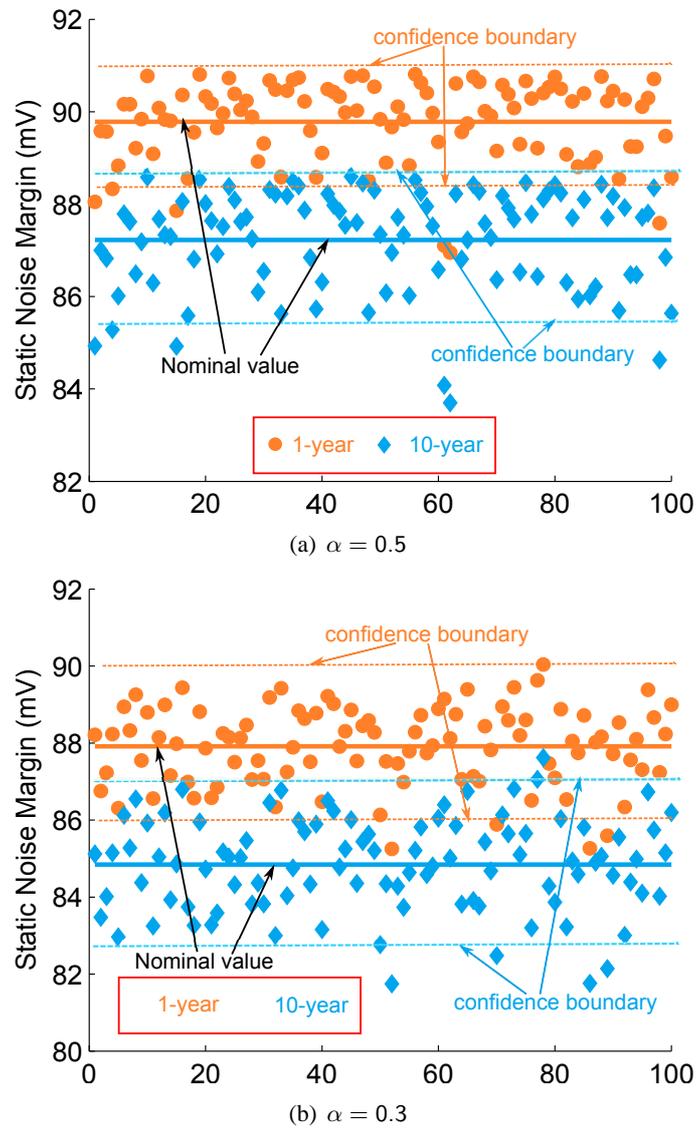


Figure 5.15: SNM fluctuations under normally distributed NBTI duty cycles α cases. For both cases the deviation $\sigma(\alpha)$ are set to be $0.2\mu(\alpha)$.

and cycle time to balance the duty-cycles between the two internal nodes of the SRAM cell, we argue that the asymmetric degradation of the two nodes can be compensated by slightly increasing the *VBPG* bias.

To demonstrate our argument, we generated two sets of normally distributed α ratios to simulate the non-uniform NBTI duty-cycles in the SRAM cells. In one set of the ratios, the mean value of α is set to be 0.5, simulating the symmetric (i.e., $\alpha_L = \alpha_R$) case of workload for the two nodes of the SRAM cell; while in the other set of ratios, the mean value of α is set to be 0.3, simulating the asymmetric case of workload for the two nodes. The real distribution of α ratios is out of the scope of this work, but it should be close to the symmetric case. This holds true for example in the case of general purpose processors, when during the application execution, the probabilities of "0"/"1" occurring in a bit cell are very fast becoming equal, and the probability that one cell bit is always "0" or "1" is extremely small.

The SNM degradation results are presented in Figure 5.15. As we can observe in the figure, the deviation of the symmetric workload case is indeed smaller than the one corresponding to the asymmetric workload case. However, the spread of cell SNM degradation due to α variations for both cases is relatively small (about 2%) compared with the SNM. Further, Figure 5.16 presents the corresponding required *VBPG* bias for these two non-uniform α -ratio cases. The blue and orange area in Figure 5.16 stands for the required *VBPG* increment to cover $3\text{-}\sigma$ SNM variation induced by ΔV_{th} and the α -ratio randomness in the SRAM array, respectively. As one can observe in the figure, the *VBPG* increment induced by α -ratio randomness is smaller than the one induced by ΔV_{th} variations for both symmetric and asymmetric workload cases. Comparing the symmetric and asymmetric workload cases, one can find out that the required *VBPG* value in the asymmetric case is just slightly larger than the one in the symmetric case. Even though the stress ratio in all the cells can be perfectly balanced, the SNM deviation induced by ΔV_{th} variations is still large. In other words, without V_{th} compensation technique, the failure rate caused by stability fluctuation will be still very large for perfectly stress-balanced SRAM array using techniques like the cell-flipping proposed in [58]. In contrast, the proposed V_{th} compensation technique in this work is able to maintain the required cell stability in the presence of ΔV_{th} variations and α -ratio randomness.

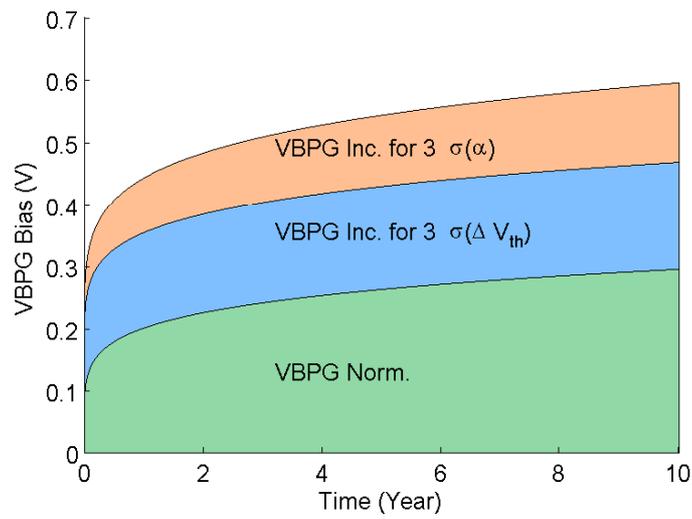
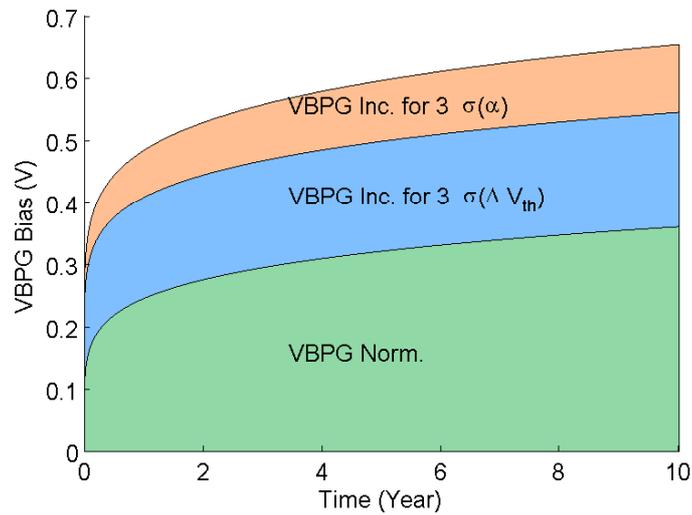
(a) $\alpha = 0.5$ (b) $\alpha = 0.3$

Figure 5.16: The required VBPG bias for non-uniform α ratios in SRAM arrays. For both cases the deviation $\sigma(\alpha)$ are set to be $0.2\mu(\alpha)$.

5.6 Conclusion

In this chapter, we investigated the influence of NBTI degradation induced variation and random process variations on the stability of the 6T-SRAM cell. Based on SPICE simulations, we thoroughly examined the contributions of ΔV_{th} variation in different transistors to the cell stability. After that, we proposed a variation mitigation technique able to maintain the SRAM cell stability within a targeted range. Our proposed approach relies on an I_{DDQ} current sensor, and on the FinFET capability to operate in the independent mode to assess the degradation level and to perform threshold voltage compensations, respectively. We evaluated the impact of our proposal on the SRAM stability by means of SPICE simulations for 20nm FinFET devices. Simulation results indicate that the proposed technique can effectively maintain the stability of an SRAM array within the desired range during its operational life under both spatial and temporal variations, hence substantially improve the performance and reliability of the system. For example, our method allows for maintaining the SNM degradation of SRAM cells to 2% of fresh device after 1 year operation, which is about 55.56% improvement when compared with the 4.5% degradation corresponding to the uncompensated case.

Mitigating the stability degradation induced by aging failure mechanisms in SRAM arrays is efficient and relatively simple, due to the simple cell structure and the regular array organization. However, the proposed self-adaptive biasing techniques might not be suitable for combinational logics because it relies on the assumption that the degradation is relatively uniformly distributed in the cell array, which is normally not the case in combinational logics. In order to achieve an effective and efficient reliability management for combinational logics, a reliability assessment technique will be presented in the next chapter.

Note. The contents of this chapter is based on the the following papers:

*Y. Wang, S. D. Cotozana, L. Fang, **Statistical Reliability Analysis of NBTI Impact on FinFET SRAMs and Mitigation Technique Using Independent-Gate Devices**, Proceedings of 2012 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH), 2012, pp. 109–115.*

*Y. Wang, S. D. Cotozana, L. Fang, **Analysis of the Impact of Spatial and Temporal Variation on the Stability of SRAM Arrays and Mitigation Technique Using Independent-Gate Devices**, accepted by Journal of Parallel and Distributed Computing (in press).*

6

Dynamic Reliability Management - Reliability Assessment

Aggressive technology scaling has led Integrated Circuits (ICs) suffer from ever-increasing wearout effects. As a consequence, Dynamic Reliability Management (DRM) becomes an essential approach to safeguard IC's lifetime reliability. Accurate and efficient reliability modeling based on low-level aging sensor measurements is the most critical part of DRM systems. This chapter presents a lifetime reliability modeling and enhancing framework, which utilizes the aging information from dedicated low-level aging sensors to extrapolate the overall IC health status. We first proposed a path delay shift model to link up the measured aging information with the actual circuit degradation. Then we introduce a Timing-Sharing Sensing (TSS) method for V_{th} -based DRM to accurately sample the dynamic activity ratio from the circuits under monitoring with limited hardware overhead. Furthermore, we propose a system reliability model utilizing aging data from sensors and investigate the estimation error caused by process variations. We demonstrate our methods by means of SPICE simulations on a representative pipeline composed by several benchmark circuits and the simulation results indicate that the proposed TSS method can significantly improve the accuracy of reliability assessment at circuit level without requiring large area overhead. Furthermore, our results indicate that the variability of sensor readings is a key issue that should be considered in the evaluation process in order to obtain accurate reliability assessment data.

6.1 Introduction

Due to the aggressive MOSFET technology scaling that took place in the past decades, reliability has been becoming a rising concern for processor designers. The increasing power densities and operating temperatures made Integrated Circuits (ICs) suffer from multiple intrinsic failure mechanisms during their servicing time. As a result, the device performance degrades gradually and consequently might lead to fails during IC's expected lifetime [72].

To combat the gloomy outlook of reliability situation, a variety of techniques have been proposed to ensure ICs' lifetime reliability. At design-time, an extra delay slack called "guardband" is normally added to the clock period to tolerate performance degradation due to wearout. However, as process and environmental variations like supply voltage and temperature are expected to become reliability threats in advanced technology nodes, one can hardly make ICs to meet their lifetime reliability specifications just by design-time approaches.

On the contrary, Dynamic Reliability Management (DRM) techniques attempt to hide the inherent pessimistic reliability landscape while maintaining the same performance and lifetime expectation. A first step towards addressing this issue called "RAMP" is proposed in [97], which is a micro-architecture level model that allows performance boosting within an acceptable reliability margin. Recent proposals introduced task scheduling and dynamic voltage and/or frequency scaling techniques to recover the performance lost to process variation [78, 107]. Some other efforts have been contributed to implement thermal management to extend processor lifetime [67, 119]. There have been also attempts to design adaptive pipelines with flexible cycle-time for different stages [109]. However, most of these existing approaches have shortcomings, e.g., they rely on a statistical modeling for reliability prediction, lacking direct health information from dedicated aging sensor therefore losing accuracy, or do not have an accurate system-level reliability assessment and prediction module to make proper use of the aging data collected from low-level sensors, thus losing efficiency and accuracy within the reliability aware optimization process.

Though many aging sensors based on different principles have been proposed [44, 46, 47, 53], however, how to efficiently utilize the collected aging information for reliability optimization is still an open issue. Accurate modeling and assessment of system reliability is critical to DRM systems, thus a quantitative reliability model, taking full advantage of the deployed on-chip aging sensors is at premium. Addressing this issue, in this chapter, we make

the following contributions:

- We develop a circuit level delay shift model, which makes use of the collected information from dedicated aging sensors.
- We propose a Timing-Sharing Sensing (TSS) method for existing V_{th} -based DRM systems meant to combat NBTI-induced wearout in ICs. The TSS method can accurately sample the activity ratio from the circuits under monitoring with a reduced number of aging sensors, which is crucial for the practical implementation of any DRM system.
- We introduce n -Time-To-Failure (n TTF) based "virtual age" concept and utilize it as the system reliability metric, instead of the conventional Mean-Time-To-Failure (MTTF). Compared with MTTF, the virtual age metric sets a more reasonable degradation budget for the system, thus reduces the area and power consumption overhead.
- We evaluate our proposal by means of simulations. We utilized multiple benchmark circuits and SPICE simulation results indicate that the proposed circuit reliability model and TSS method can achieve accurate reliability assessment with significantly reduced area and power overhead when compared with existing schemes equivalent state of the art approaches.

The rest of the chapter is organized as follows. Section 6.2 introduces the required background. Section 6.3 presents the path delay shift model that makes use of the collected information from dedicated aging sensors. Section 6.4 introduces the TSS scheme based reliability assessment for V_{th} variation driven DRM systems. Section 5 describe the experimental methodology and simulation results. Section 6 concludes this chapter.

6.2 Conventional DRM Scheme with V_{th} -Based Sensor

A typical DRM scheme with aging sensors is illustrated in Figure 6.1. Most aging sensors are designed to quantitatively measure a physical parameter which is correlated with the aging status of a device. A typical example are the vastly investigated threshold voltage (V_{th}) based sensors tailored to assess the consequence of the NBTI failure mechanism [44, 53]. However, previous works did not answer the question: how would high level DRM systems utilize the low-level aging data collected from aging sensors to achieve a better optimization

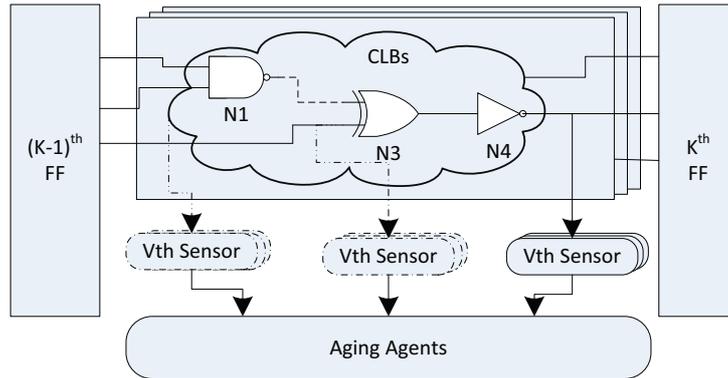


Figure 6.1: Typical DRM scheme with critical path under monitoring. Multiple sensors are required to monitor a single critical path.

for reliability, in the sense of accuracy and efficiency. Though the measured physical parameter can be used as a proxy stub of the aging status of the particular Device-Under-Test (DUT), how would it reflect the overall health status of the system, especially how to use it in a cost function for high-level reliability optimization is still under investigation.

Addressing the issues discussed above, we propose a system-level reliability modeling framework, which utilizes aging information from low-level aging sensors to achieve accurate and efficient reliability optimizations. The aging sensors we considered in this chapter are V_{th} -based aging sensors, which are associated to measure degradation from a single device but our approach is more general, in principle, can also be applied in conjunction with other sensor types.

6.3 Delay Shift Due to Aging

NBTI is considered as the most dominant wearout mechanism in current and future technology nodes [72] and it is mainly experienced in PMOS with the channel in inversion. As we have already discussed in Chapter 3, NBTI is modeled under Reaction-Diffusion (RD) theory, which gives a power-law time evolution of V_{th} degradation. A long-term V_{th} shift under dynamic NBTI stress is given by [16]:

$$\Delta V_{th} = \left(\frac{\sqrt{K_v^2 \alpha_s T_{clk}}}{1 - \beta_t^{1/2n}} \right)^{2n}, \quad (6.1)$$

where $n = 1/6$ is the power-law time constant, K_v is the accelerating coefficient (including temperature and electrical field), T_{clk} is clock period, α_s ($0.1 \leq \alpha_s \leq 0.9$) is the stress probability of NBTI, i.e., NBTI duty-cycle, and β_t is a coefficient reflecting the recovery effect of NBTI.

Under high frequency dynamic stress, Eq.(6.1) can be approximated as [16]:

$$\Delta V_{th} = A \cdot (\alpha_s t)^n, \quad (6.2)$$

which gives the dependence of V_{th} on the duty cycle α_s . As α_s can hardly be determined at design time, V_{th} sensors are utilized to assess the aging status of the circuit critical paths at runtime, in order to perform an efficient reliability-aware resource management.

According to the alpha-law model [93], the propagation delay of a CMOS device can be estimated as:

$$t_{pd} = \frac{C_L V_{DD}}{\beta(V_{DD} - V_{th})^\alpha}. \quad (6.3)$$

where α and β are fitting parameters, C_L is the load capacitance, V_{DD} is the power supply voltage, and V_{th} is the threshold voltage. Hence, the first order of delay shift due to V_{th} degradation can be expressed as:

$$\Delta t_{pd} = \frac{\partial t_{pd}}{\partial V_{th}} \Delta V_{th} = \frac{\alpha t_{pd0} \Delta V_{th}}{V_{DD} - V_{th0}} = K_t \Delta V_{th} \cdot t_{pd0}, \quad (6.4)$$

which suggests that the fraction of delay shift is proportional to the V_{th} degradation value, K_t being the correlation efficient. V_{th} aging sensors are designed to replicate the aging stress of a device in the Critical Path (CP), such that the aging status of CPs can be estimated from the aging status of sensors. The delay shift of a logic gate in the CP is correlated to the V_{th} degradation of the aging sensor through environmental factors, including signal probability (SP), environmental factors V_{DD} and temperature T .

Eq. (6.4) gives the delay shift of a single PMOS device, which can be attributed the NBTI stress induced reduction of the maximum drain current. In the case of gate-level delay shift, the gate logic structure has to be taken into consideration as well. As illustrated in Figure 6.2(a), the 2-input NAND gate has two pull-up paths through PMOS A and B, respectively. So the delay shift contribution of an n -input NAND gate ΔD_{NAND} to a specific logic path should be the delay shift of the corresponding aged PMOS in the NAND gate, i.e.,

$$\Delta D_{NAND} = \Delta t_{pd,i}, i \in \{1, 2, \dots, n\}, \quad (6.5)$$

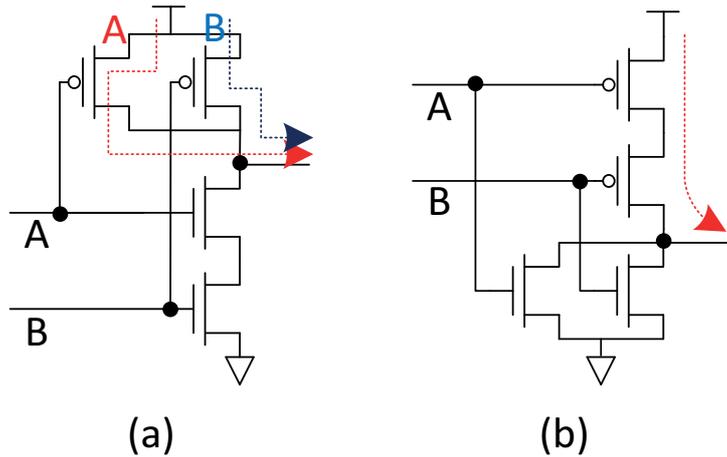


Figure 6.2: CMOS Logic Gates with Pull-up Path(s): (a) 2-NAND; (b) 2-NOR.

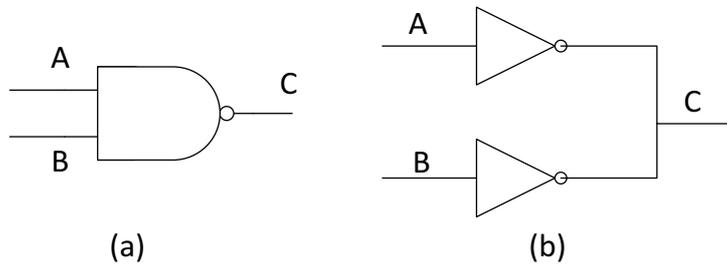


Figure 6.3: Using 2 inverters to substitute the 2-input NAND gate.

where $t_{pd,i}$ is the propagation delay through the i^{th} PMOS device in the NAND pull-up network.

In other words, only one PMOS device's degradation contributes to the rising time shift in NANDs. However, Eq. (6.5) requires an identification process on the PMOS device when utilizing it to calculate the delay shift, which is not convenient in practice. In order to simplify the calculation, we can use inverters to substitute an n -input NAND gate, as demonstrated in Figure 6.3. After the substitution, the n -input NAND gate splits up into n independent inverters, and the inverter delay shift represents the delay shift of the n -input NAND delay shift in the corresponding circuit path.

For NOR gate, as illustrated in Figure 6.2(b), the degradation of all the PMOS devices contribute to the degradation of the gate performance. The NOR delay

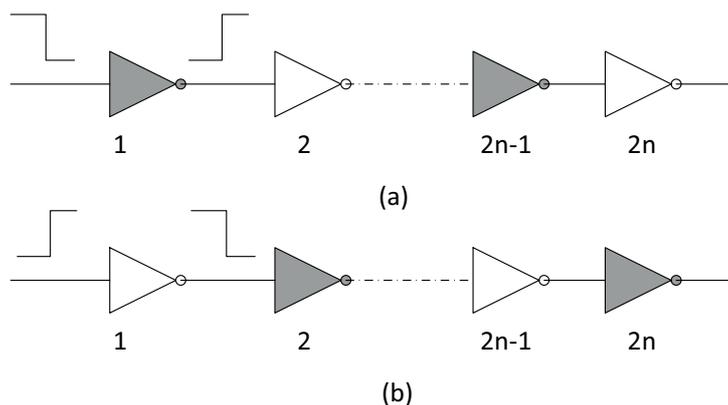


Figure 6.4: Inverter chain with: (a) high to low transition; (b) low to high transition. Only the degradation of inverters with shadow contributes to the NBTI stress induced delay shift.

shift can be written as:

$$\Delta D_{NOR} = \sum_{i=1}^n \Delta t_{pd,i}. \quad (6.6)$$

One can easily deduce from Eq. (6.6), NOR gates are more prone to the NBTI stress induced aging. However, the probability that an NOR gate contributes to the delay shift of an entire circuit might be quite lower than the one corresponding to an NAND gate. This can be justified by the fact that all the NOR gate inputs have to be "0" to turn on the pull-up path, while one "0" input is enough to turn on a NAND pull-up path.

It is also worth to notice that not every gate in a circuit path contributes to the delay shift even though every gate is aged due to NBTI stress. This can be illustrated on the inverter chain shown in Figure 6.4. Based on the assumption that NBTI is negligible in NMOS devices, only the inverters which pull the output node up will contribute to the path delay shift. Other failure mechanisms like PBTI may damage the pull-down network as well, but the fact that different logic stages contribute to the path delay shift due to degradation in either pull-up or pull-down network still holds true.

The delay shift of a CP $\Delta\tau_{cp}$ can be approximated as the sum of the Pull-Up-Gate (PUG, i.e., the gates with at least one pull-up path is connected) delay

shifts inside it, which can be written as:

$$\Delta\tau_{cp} = \sum_{i \in PUGs}^l \Delta D_i = k_\tau \sum_{i=1}^h \Delta V_{th,i}, \quad (6.7)$$

where l is the CP depth, ΔD_i is the delay shift of a single logic gate, and h equals the number of PMOS devices in the PUGs. Notice that only a part of the logic gates (e.g., odd-/even-stage gates in an inverter-chain) in the CP are in the PUGs during an input signal transition. And k_τ is a V_{th} dependence coefficient, which can be obtained from Eq. (6.4) and be expressed as follows

$$k_\tau = \frac{\alpha t_{pd0}}{V_{DD} - V_{th0}} \cdot \Delta V_{th} = K_t \cdot t_{pd0}. \quad (6.8)$$

Eq. (6.7) suggests that the NBTI-induced delay shift of a critical path is a sum of the active PUG delay shifts. Since the active PUGs rely on the input signal patterns and it is not feasible to deploy an aging sensor for each and every input signal in the critical paths due to area and power overhead limitations, an efficient aging assessment scheme is highly desirable in practical DRM implementations. Addressing on this problem, we introduce a time-sharing aging assessment scheme in the next section.

6.4 Time-Sharing Sensing Scheme for Aging Assessment

In order to obtain an accurate estimation on $\Delta\tau_{cp}$, the aforementioned environmental correlating factors have to be carefully examined. Generally, the aging sensor are deployed close enough to the logic gates in the CP, in order to sample a same temperature and supply voltage operating conditions with the devices/circuits. Due to its very operating principle, i.e., the stress measurement is done in an indirect way by replicating the logic input signals into a device in the sensor (See Section 6.2), one V_{th} sensor can just sample the duty-cycle from one input node of a logic gate. Due to obvious practical reasons, i.e., prohibitive area and power consumption overheads, we cannot deploy an aging sensor for each and every input signal of the CP logic gates, we need to find a way to estimate the NBTI duty-cycle, i.e., Signal Probability (SP) for the logic stages with information from a limited number of aging sensors.

However, to determine the dynamic SP for an n -gate CP with p -node inputs for each gate is a NP-complete problem. To simplify this problem, existing

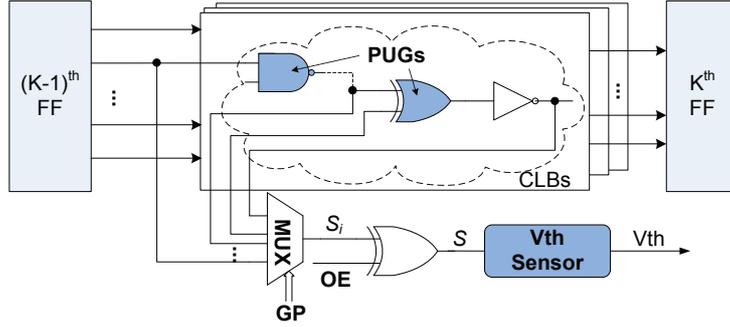


Figure 6.5: Time-Sharing Sensing Scheme for Critical Path Monitor.

DRM systems [41, 42, 97] either assume or imply a uniform SP within the associated aging sensor for all logic stages in the CP. This assumption is very convenient from the computation overhead point of view, but it's far from providing accurate results. In fact, the SP estimation difficulty comes from the temporal and spatial randomness of signals. We address this issue by proposing a Time-Sharing Stressing (TSS) scheme for the more effective utilization of the aging sensors. As illustrated in Figure 6.5, instead of sampling activity from the same node all the time, the aging sensor takes stress signals from different nodes in the CP at different moments in time. This is achieved by routing all the necessary signal from the CP to a MUX, which selects one input signal and sends it to the V_{th} sensor. The input signal selection is based on a round-robin algorithm, and it switches at every time interval Δt . This new method substantially diminishes the area overhead, and lets the sensor sample as much information as possible from the entire CP instead of one single node.

Based on the proposed scheme, the ΔV_{th} of aging sensor after N intervals can be expressed by:

$$\Delta V_{th} = \sum_{i=1}^N \Delta V_{th,i} = \sum_{i=1}^N A_i (\alpha_i \Delta t_i)^n, \quad (6.9)$$

where $\Delta V_{th,i}$ is the parameter degradation during the i^{th} time interval. Further assume $N = k \cdot h$, then the time intervals can be divided into k sub-sequences $\Delta T_j = h \cdot \Delta t$, where $j = 1, 2, \dots, k$. For the j^{th} sub-sequence, assume the sampled activity ratio of the m^{th} logic gate is the activity ratio for the entire time span, i.e., $\bar{\alpha}_{t \in T_j} \approx \alpha_{t \in (T_{j-1} + m\Delta t)}$, then according to Eq. (6.1), the total

ΔV_{th} of the m^{th} logic gate at the end of sub-sequence ΔT_j can be written as:

$$\Delta V_{th,t \in \Delta T_j} = \left(\frac{\Delta T_j}{\Delta t} \right) \cdot \Delta V_{th,t \in \Delta t_m} = h \Delta V_{th,t \in \Delta t_m}, \quad (6.10)$$

where t_m is the m^{th} interval Δt in ΔT_j . Insert Eq. (6.10) into Eq. (6.7), the delay shift of the entire CP during a large time interval ΔT_j can be estimated as:

$$\Delta \tau_{cp,t \in \Delta T_j} = h k_\tau V_{th,t \in \Delta T_j}. \quad (6.11)$$

Then, the accumulated delay shift can be estimated as:

$$\Delta \tau_{cp} = \sum_{j=1}^k \Delta \tau_{cp,t \in \Delta T_j} = h k_\tau V_{th}. \quad (6.12)$$

Eq. (6.12) suggests that by introducing the TSS scheme into a DRM system, the long-term delay shift of a critical path is proportional to the V_{th} degradation measured from the associated aging sensor. In other words, the V_{th} sensor measurement indicates the NBTI-induced degradation of the entire CP in TSS scheme, instead of the degradation of one single device in the conventional configuration, which is a significant improvement in terms of area and power efficiency.

6.5 Delay Shift Calibration under Process Variations

The path delay shift model can just tell the aging status of one CP, based on information from the associated aging sensor. In order to obtain the system-level reliability status, multiple (but with a limited number) aging sensors have to be deployed onto the chip. However, sensor readings can be affected by the commonly existing process variations and environmental non-uniformity. On the other hand, the pre-selected CP under monitoring at design time might not be the most aging-prone CP during run-time. Therefore, how to determine the system-level reliability from aging information collected from multiple sensors is still a challenge, which we address in the remainder of this section.

Apart from input signal activity and environmental factors like temperature and V_{DD} , the aging progress interacts with the presence of intrinsic variability caused by the fabrication process variations as well. The measured data from the V_{th} sensors are affected by the process variations, too. To minimize the reliability estimation error, the reading data from multiple sensors must be

calibrated by taking into consideration the fabricated process parameter variations.

In the presence of global and local process variations, V_{th} at time 0 can be expressed as:

$$V_{th0} = V_{th,normal} + \Delta V_{thg} + \Delta V_{thl}, \quad (6.13)$$

where $V_{th,normal}$ is the nominal threshold voltage, ΔV_{thg} and ΔV_{thl} are the V_{th} variations caused by global and local variations, respectively. If we denote $\Delta V_{th_{nbt}^{pv}}$ as the influence of variations on the NBTI degradation, we can determine its value from the NBTI-induced V_{th} shift as:

$$\Delta V_{th_{nbt}^{pv}} = \frac{\partial \Delta V_{th_{nbt}}}{\partial V_{th}} (\Delta V_{thg} + \Delta V_{thl}).$$

From Eq.(6.1) we can get:

$$\Delta V_{th_{nbt}^{pv}} = -\frac{2n(\Delta V_{thg} + \Delta V_{thl})}{V_{gs} - V_{th,normal}} \cdot \Delta V_{th_{nbt}0}, \quad (6.14)$$

where $\Delta V_{th_{nbt}0}$ is the NBTI-induced V_{th} degradation without process variation. By inserting the above equation into Equation (6.3) and consider that for PMOS $V_{gs} - V_{thp,normal} = -(V_{dd} - V_{th,normal})$, we can estimate the gate delay shift under process variations as:

$$\Delta \tau_{s,pv} = (1 + \kappa(\Delta V_{thg} + \Delta V_{thl})) \cdot \Delta \tau_{s0}, \quad (6.15)$$

where $\Delta \tau_{s0}$ is the gate delay without process variation, and $\kappa \propto 1/(V_{dd} - V_{th})$ is a sensitivity coefficient. Equation (6.15) simply indicates that the gate delay shift is proportional to the V_{th} shift due to process variations. Since the global variations have the same amount of influence on both logic gates and aging sensors in the same die, the estimation error on delay shift by using sensor data is determined by the local variations, which can be expressed as:

$$\delta \tau_s = \kappa(\Delta V_{thl,gate} - \Delta V_{thl,sensor}) \cdot \Delta \tau_{s0}. \quad (6.16)$$

Therefore, the total estimation error in a path delay shift assessment is:

$$\delta \tau_{cp} = \kappa \sum_{i=1}^n (\Delta V_{thli,gate} - \Delta V_{thli,sensor}) \cdot \Delta \tau_{s0}. \quad (6.17)$$

Eq. (6.17) suggests that the total path delay shift estimation error is determined by two factors: (i) the local process variation of the gates in the path, and (ii) the local process of the corresponding aging sensor that is associated to the gate. The effects of these two factors can be cancelled by each other if we carefully design the aging sensor, making its local variation match the one in the gate that it is associated to.

6.6 Reliability Metric for DRM System

In Chapter 3 we have introduced a "lifetime" definition for CMOS devices, which is set to be the point that some device parameter (e.g., V_{th}) reaches a critical point, for instance, 10%. This definition is useful for a single device, however, it might not be suitable to be utilized in a system. For example, it is possible that the V_{th} degradation exceeds the critical point (e.g., 10%) for a device in a CP, but the total path delay shift might be still within the range required by the circuit normal functionality. To address this issue, we introduce a new reliability metric for DRM system in this section.

6.6.1 MTTF-Based Lifetime Definition

Most of the conventional DRM systems derive a Mean-Time-To-Failure (MTTF) model as the objective function of the system-level reliability optimization process. MTTF is calculated from the low-level physics-based degradation model [98], thus no aging information from sensors is required. However, without aging sensors support, the conventional DRM schemes have no ability to look inside into the real stress profile experienced by the circuits. This approach leads to a "blind" reliability optimization process thus the estimation/prediction accuracy falls dramatically.

MTTF is widely accepted as a reliability metric in industry, however, just as P. Ramachandran et al. point out in [89], MTTF loses too much information in averaging, which makes it an inaccurate indication for lifetime reliability, especially for the early life reliability prediction.

To illustrate this problem, we can take Weibull distribution [87] as an example. Weibull distribution is normally utilized to express the failure rate of a "weakest-link" system, in which the entire system fails when if one component in the system fails. The Probability Density Function (PDF), $f(t)$, of Weibull distribution can be expressed as

$$f(t) = \frac{\beta}{t} \left(\frac{t}{\alpha}\right)^{\beta} \cdot e^{-\left(\frac{t}{\alpha}\right)^{\beta}}, \quad (6.18)$$

and its Cumulative Distribution Function (CDF), $F(t)$, is:

$$F(t) = \int_0^t f(t) dt = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}}, \quad (6.19)$$

where α is the scale parameter and β is the shape parameter. Weibull's MTTF

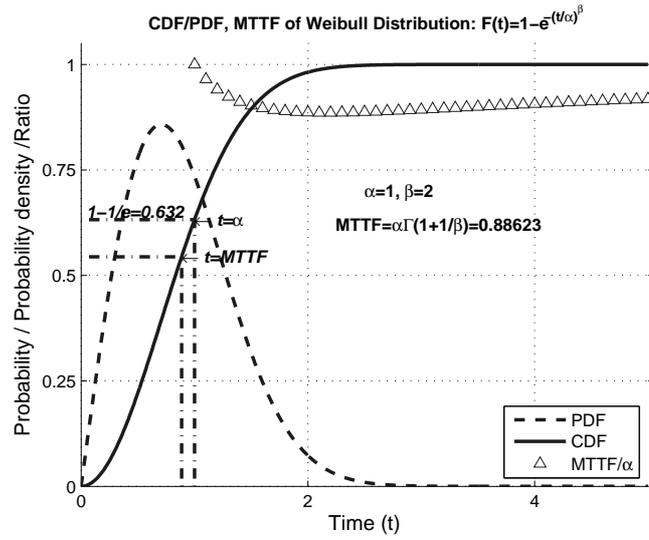


Figure 6.6: Relationship Between MTTF and the Weibull Distribution's Parameter.

is:

$$MTTF = \alpha \Gamma \left(1 + \frac{1}{\beta} \right), \quad (6.20)$$

where $\Gamma(n)$ is the Gamma-function.

α is also called the "characteristic parameter" because it indicates the time that the accumulated failure rate reaches $1 - 1/e$ (63.2%), which is far beyond the normal operational expectation (10%). For a degrading system ($\beta \geq 1$), the relationship between MTTF and scale parameter α in Weibull Distribution is graphically depicted in Figure 6.6. As depicted in the figure, the $MTTF/\alpha$ ratios (the triangles in the figure) becomes nearly constant when α is larger than 1.5. In other words, the MTTF value of Weibull distribution is proportional to the characteristic parameter α . As a result, the failure rate at the time $t = MTTF = k \cdot \alpha$ is far beyond the normal operating conditions as well. The reliability optimization approaches based on MTTF is too pessimistic and consequently leads to a over design on reliability.

6.6.2 nTTF-Based Lifetime Definition

To overcome the MTTF shortcomings the n -Time-To-Failure ($nTTF$) can be utilized instead, which is defined as the time when the device/system parameter degrades by a critical portion (n) such that the device/system no longer

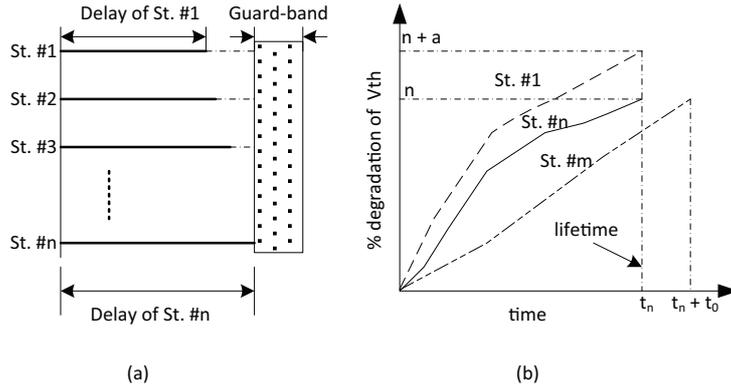


Figure 6.7: Illustration of lifetime definition:(a) Logic paths in a pipeline may be not designed with equal delays. A delay guard-band is typically added to combat for aging-induced performance degradation.(b) Lifetime is determined by the path which first eats up all the guard-band. Some paths like St.#1 can have more severe degradation but still meet reliability specification because they have more headroom of delay at time 0.

properly functions. Formally $nTTF$ can be defined as follows

$$nTTF = f^{-1}(\Delta P = n \cdot P_0), \quad (6.21)$$

where P is the degrading parameter and $\Delta P = f(t)$ is the parameter degradation function.

According to Eq. (6.21), the lifetime of a circuit can be defined as its critical path delay shift τ_{cp} induced by aging mechanisms reaches a critical portion of the clock period, which can be expressed as:

$$nTTF(t) = f^{-1}(\Delta \tau = n \cdot D_0) = f^{-1}(nD_0 - \tau_{cp}(t)). \quad (6.22)$$

where D_0 is the clock period in the processor.

Logic paths in a pipeline may be not designed with equal delays. A delay guard-band is typically added to combat for aging-induced performance degradation. This is illustrated in Figure 6.7. Lifetime is determined by the path which first eats up all the guard-band. Some paths like St.#1 can have more severe degradation but still meet reliability specification because they have more headroom of delay at time 0. With the lifetime definition in Eq. (6.22), the various delay headroom of different CPs for aging is taken into account naturally. From Eq. (6.2), Eq. (6.3) and Eq. (6.7) we can estimate the $nTTF(t)$ as:

$$\overline{nTTF}(t) = \frac{1}{\bar{\alpha}_s} \cdot \left(\frac{nD_0 - k_r \Delta V_{ths}(t)}{\bar{A}(V, T)} \right)^{\frac{1}{n}}, \quad (6.23)$$

where $\bar{\alpha}$ is an estimated average circuit activity ratio $\Delta V_{ths}(t)$ is the sensor reading at time t , n is the critical portion of the clock period D_0 , $\bar{A}(V, T)$ is the average temperature and voltage accelerator from Eq. (6.2), and k_τ is a V_{th} dependence coefficient defined in Eq. (6.8). We note that the remaining lifetime estimation (i.e., \overline{nTTF}) for a circuit due to wearout should be always associated with an estimation of the workload the circuit will carry on in its future life. Also notice that the temperature and voltage accelerator $A(V, T)$ is also time-varying parameter for a long-term estimation. Near-term or short-term estimation on these parameters is possible with certain assumptions and algorithms, but applying those methods to a MTTF-term optimization will lose its accuracy and practicability because of the validity of the assumptions.

The lifetime definition of a Functional Unit (FU, assume all the combinational paths of it are in a same pipeline stage) can be defined as the minimum $nTTF(t)$ over all the logical paths within the FU at time t . To obtain the overall lifetime of processor, with considering the structural redundancy which commonly exists in modern processor cores, we apply the MIN-MAX method in [98] to $nTTF$ instead of MTTF, as proposed in the original work.

6.6.3 Virtual-Age Definition for Multiple Failure Mechanisms

The $nTTF$ definition introduced in the previous section is based on a single failure mechanism only. However, multiple failure mechanisms exist simultaneously in the circuits. To take this scenario into consideration in the model, we introduce a "virtual age" concept and utilize it as the system reliability metric for multiple failure mechanisms in this section.

The term "virtual age" is originally defined as the corresponding equivalent age of a repairable item when a repair is imperfect [32]. But a typical processor chip is a non-repairable system, thus we define the "virtual age" as the reliability status relative to a standard baseline.

Let us assume two comparable items working in two environments: one in an identical continuous baseline environment where temperature (or other concerned factors) is constant, and the other in stochastic severe real working condition. The virtual age can be calculated by referring the reliability in the severe environment equivalently to the one in the baseline environment, where has a following relationship

$$R_s(t) = R_b(t_s), \quad (6.24)$$

where $R_x(t) = 1 - F(t)$ ($x \in \{s, t\}$) is the cumulative reliability rate function

(also referred as "survival rate function"), and $R_s(t)$ and $R_b(t)$ are the reliability functions in the severe environment and in the baseline environment, respectively. Then, the virtual age in the severe environment can be defined by referring a equivalent reliability rate in the baseline environment, which can be expressed as

$$t_s = V(t) = R_b^{-1}(R_s(t)). \quad (6.25)$$

On a discrete case, assume at every interval, system has a uniform reliability rate CDF, then the following equation can be obtained at time t_i :

$$R_s(t_i) = R_b(t_{s,i}). \quad (6.26)$$

After a time interval Δt , we can obtain:

$$R_s(t_i + \Delta t) = R_b(t_{s,i+1}) = R_b(t_{s,i} + \Delta t_{s,i}), \quad (6.27)$$

where $\Delta t_{s,i} > 0$ is the virtual age increased during this time interval Δt . According to the CDF definition, $R(t) = Pr[T > t]$, where T is the random variable, the probability that the survival rate after a short interval Δt at time t follows the conditional probability rule, which can be expressed as

$$\begin{aligned} R_{t_x}(\Delta t|t_x) &= Pr[T > t_x + \Delta t | T > t_x] \\ &= \frac{Pr[T > t_x + \Delta t]}{Pr[T > t_x]} \\ &= \frac{R_{t_x}(t_x + \Delta t)}{R_{t_x}(t_x)}, \end{aligned} \quad (6.28)$$

where R_{t_x} is the CDF in the time span $[t_x, t_x + \Delta t)$. Substituting Eq. (6.26) and Eq. (6.28) into Eq. (6.27), we can get

$$\Delta t_{s,i} = R_b^{-1}(R_s(\Delta t_i)). \quad (6.29)$$

Eq. (6.29) presents the virtual age increment after a short time interval Δt at time t_i . If we sum all the virtual age increment up from time 0, we can obtain the virtual age as expressed in the following equation:

$$\begin{aligned} V(t_n) &= t_0 + \sum_{i=0}^{n-1} R_b^{-1}(R_{s,i}(\Delta t_i)) \\ &= V(t_{n-1}) + R_b^{-1}(R_{s,n}(\Delta t_n)), \end{aligned} \quad (6.30)$$

where t_0 is the initial virtual age in case system's initial reliability is not 100%.

Generally, multiple failure mechanisms are experienced simultaneously in a system. These simultaneous failures can be modeled by a competing risky model, where system fails when any failure happens. The reliability CDF can be expressed as

$$R(t) = \prod_{i=1}^n R_i(t), \quad (6.31)$$

and then system's virtual age can be written as:

$$V(t_n) = V(t_{n-1}) + R_b^{-1} \left(\prod_{j=1}^m R_{s,jn}(\Delta t_n) \right). \quad (6.32)$$

Eq. (6.32) presents a universal method to sum up the aging effects induced by multiple different failure mechanisms, even though they might have influence on different device parameters. The increasing speed of system virtual age stands for the degradation severity in the system, and the virtual age value can be used as the cost function in the reliability optimization process.

6.7 Experimental Results

In order to validate the proposed reliability modeling framework, simulations are conducted for a set of ISCAS85 and ISCAS89 benchmark circuits. A representative 5-stage pipeline is formed by the C432, C499, C880, C1908, and C3540 circuits, with each circuit as a pipeline stage. A PTM [5] 32nm technology library is used for SPICE simulation. The technology library consists of 5 different cells: INVERTER, 2-input NAND, 3-input NAND, 2-input NOR, and 3-input NOR. The NBTI induced degradation is simulated by threshold voltage degradation, as indicated by Eq. (6.1). The simulations for SPICE are performed in HSPICE at an accelerated condition, i.e., $T = 120^\circ\text{C}$, which results in an equivalent lifetime of 10-year operation at normal conditions.

In order to investigate the proposed delay shift estimation model, we feed all the benchmark circuits with 3 groups of input signals, and each group of signal is set to be with a 30%, 50% and 70% constant activity ratio, respectively. The activity ratios at the internal nodes in the benchmarks are dumped through circuit behavioural simulations, and then the ratios are utilized in the model to calculate the path delay shifts. In such a way, the activity ratios in the model are ideal, thus the estimation error is introduced by the model itself only. For comparison, the "real" (or "ideal") path delay shifts are extracted by SPICE simulations after the aforementioned accelerated aging stressing process.

Table 6.1: Delay Shift Estimation for ISCAS85, 89 Circuits.

Name	# Gates	$\alpha = 30\%$			$\alpha = 50\%$			$\alpha = 70\%$		
		$\Delta D(\%)$	Est.(%)	Err.(%)	$\Delta D(\%)$	Est.(%)	Err.(%)	$\Delta D(\%)$	Est.(%)	Err.(%)
C432	303	14.35	17.77	3.42	14.78	16.87	2.09	14.52	16.59	2.07
C499	202	13.62	18.68	5.06	13.93	17.95	4.02	13.07	19.31	6.24
C880	383	14.35	16.54	2.19	15.02	17.06	2.04	14.13	17.34	3.21
C1908	817	15.89	21.34	5.45	14.79	19.28	4.49	15.37	16.86	1.49
C3540	1669	13.31	19.64	6.38	14.12	18.32	4.20	13.67	18.78	5.11

Table 6.2: Delay Shift TSS Based Estimation for ISCAS85, 89 Circuits

Name	Estimation Error by TSS Scheme relative to D_0 (%)							
	$\Delta t = 1000$ clk		$\Delta t = 2500$ clk		$\Delta t = 5000$ clk		$\Delta t = 10000$ clk	
	Cst.	Dvt.	Cst.	Dvt.	Cst.	Dvt.	Cst.	Dvt.
C432	2.34	3.28	2.31	2.73	2.26	2.77	2.19	2.32
C499	4.73	6.17	4.45	5.09	4.27	3.46	4.11	3.89
C880	2.89	2.21	2.64	3.17	2.37	2.55	2.17	2.25
C1908	4.18	3.96	4.73	8.07	4.65	4.12	4.53	4.79
C3540	4.46	2.95	4.37	3.51	4.28	3.29	4.25	5.03

The simulation results are summarized in Table 6.1. The first and second columns list the circuit's name and number of gates, respectively. The rest columns are the percentage of delay shift measure from SPICE simulation (ΔD columns, i.e., the "real" delay shifts percentages), and the delay shift estimated by the proposed model (*Est.* columns) for $\alpha = 30\%$, 50% , and 70% , respectively. The differences between these two methods are presented in the *Err.* columns. As we can deduce from the *Err.* columns, the proposed model overestimates the delay shift. The reason for that might be related to the fact that our model sums up the degradation of each PMOS device in NOR gates, which is a pessimistic estimation of the real conditions. This error can be reduced by introducing a more realistic model for the stacking effect, however, this complicates the aging sensor sampling scheme and the online reliability assessment algorithm. However, this pessimistic estimation is still in an acceptable margin and can be relaxed by allowing a certain level reliability constraint violations in practice.

After that, we further investigate the proposed TSS scheme based on the previous experimental set-ups. In this time, the activity ratios for the circuit input signals are set to be constant 50% and $50\% \pm 10\%$ random deviation. the activity ratios are extracted by the proposed TSS scheme instead of that obtained by simulations. With the "sampled" activity ratios, the path delay shifts are calculated. In order to investigate the influence of the sampling period on the estimation result accuracy, four sampling periods (Δt) are simulated, which are, namely, $\Delta t = 1000$, 2500 , 5000 , and 10000 cycles.

Table 6.2 presents the absolute errors of the proposed TSS scheme for delay shift estimation. The *Cst.* columns include the results for a constant activity ratio ($\alpha = 50\%$) signal, and the *Dvt.* columns include the results for variable activity ratio signals. In order to obtain a stable TSS error estimation, the *Dvt.* columns are reported as the average value of 1000 runs with the specified con-

ditions. First, by comparing the *Cst.* columns with the *Err.* column whose activity ratio is 50% in Table 6.1, we can clearly observe that as the sampling time interval (i.e., Δt) increases, the difference between the *Cst.* column and the *Err.* column decreases. In other words, as the sampling interval increases, the accuracy of the TSS scheme assessment increases. However, the relationship is lessened by the randomness of the signal activity ratios. This can be confirmed by comparing the *Dvt.* columns with the *Err.* column ($\alpha = 50\%$) in Table 6.1: after introducing a deviation to the activity ratio, the estimation error based on the TSS scheme increases. However, the absolute error introduced by the TSS remains in an acceptable margin for all the cases ($\leq 5.03\%$).

6.8 Conclusion

In this chapter we presented a lifetime reliability modeling and enhancing framework, which investigates how to utilize the aging information from dedicated low-level aging sensors to extrapolate the overall health status. We first proposed a path delay shift model, which utilizes the collected aging information from low-level sensors to estimate the circuit path delay shift induced by failure mechanisms, like Negative Bias Temperature Instability (NBTI), and so on. Then we propose a Timing-Sharing Sensing (TSS) scheme for V_{th} -based DRM systems to assess the dynamic reliability status from the circuits under monitoring. With the proposed TSS scheme, one aging sensor can be utilized to monitor multiple paths in the circuits, which can significantly reduce the required aging sensor numbers in practical DRM implementations. SPICE simulation results have indicated that the proposed path delay shift estimation model and TSS scheme can predict/assess the circuit performance degradation, i.e., circuit path delay shift, with an acceptable accuracy margin no more than 5.03%.

Note. The contents of this chapter is based on the the following papers:

Y. Wang, S. D. Cotofana, L. Fang, A Novel Virtual Age Reliability Model for Time-to-Failure Prediction, IEEE International Integrated Reliability Workshop Final Report (IIRW), 2010, pp. 109–115.

Y. Wang, S. D. Cotofana, L. Fang, Lifetime Reliability Assessment with Aging Information from Low-Level Sensors, in Proceedings of Great Lakes Symposium on VLSI (GLSVLSI), 2013, pp.339–340.

7

Conclusions and Future Work

In this dissertation, we have presented a Dynamic Reliability Management (DRM) framework which relies on aging sensor based reliability assessment. Our approach attempts to combat aging degradation in circuits and opposite to most existing proposals, it uses extracted reliability information from online aging sensors to perform quantitative reliability management.

7.1 Summary

To give readers a general picture on the gloomy reliability landscape in nanometer CMOS technologies, in Chapter 2 we presented an introduction of the major aging failure mechanisms for background purpose. The physical fundamentals of the CMOS Integrated Circuit (IC) failure mechanisms were presented. Furthermore, the most popular physical models of the major failure mechanisms were discussed. Subsequently, we presented the state of the art in reliability aware computation on combating aging degradation. The overview of the reliability-aware computation techniques covers a wide range of research areas, from architecture level design to dynamic task scheduling and resource allocation. Most of the existing reliability aware computing platforms did not employ dedicated aging sensor to extract degradation information from circuits. As an alternative, we proposed a quantitative reliability management framework in this dissertation. As a key component of quantitative DRM systems, the existing aging sensors designs and their shortcomings were reviewed in this chapter as well.

In order to keep on the scaling trend in CMOS technologies, novel devices with new geometrical structures and/or materials are introduced in recent years. Among other novel devices, FinFET is one of the most promising device to substitute the current planar devices. In Chapter 3 we investigated the reliabil-

ity features of double-gate and triple-gate FinFET devices. In this chapter, we have made the following major contributions:

- A unified degradation model for Negative Bias Temperature Instability (NBTI) and Hot Carrier Injection (HCI) in double-gate and triple-gate FinFET devices was proposed. The proposed model unifies the NBTI and HCI degradations, which simplifies the simulation complexity thus makes it suitable for utilization in circuit simulation.
- The reduced dimension effects of nanoscale devices on the NBTI and HCI induced degradations were thoroughly investigated.
- A device performance degradation model was introduced, which is able to capture and predict the aging process due to NBTI and HCI inside FinFET based ICs.

The simulation results have suggested that our model is accurate and computationally efficient, which makes it potentially applicable for lifetime reliability management schemes to be included in reliability-aware architectures.

To extract reliability information from circuits, we proposed two kinds of aging sensor designs in Chapter 4, based on threshold voltage (V_{th}) and power supply current (I_{DD}) measurement, respectively. In this chapter, we have made the following major contributions:

- A new DRM framework that relies on quantitative degradation measurements was introduced. Our approach can quantitatively analyze the system aging status, which gives a deeper insight on the degradation progress and eventually leads to a better optimization for reliability.
- Novel V_{th} based aging sensors for NBTI and HCI degradation measurement were proposed with good Process, Voltage, and Temperature (PVT) variation tolerance.
- A Novel I_{DD} based aging sensor were proposed with the capability of directly measuring the actual IC aging status caused by the amalgamated effects of NBTI and HCI failure mechanisms.

The simulation results have indicated that the proposed V_{th} based aging sensors have good PVT-variation tolerance, which outperforms the state of the art approaches and that the I_{DD} based aging sensor is able to accurately and effectively capture the amalgamated effects of NBTI and HCI failure mechanisms.

In order to keep on the industry scaling-down trend, novel devices and structures were proposed as potential candidates to replace the conventional planar MOSFET device. In Chapter 5 we investigated the influence of NBTI degradation and random process variations on the stability of the FinFET based 6T-SRAM cell. In this chapter, we have made the following major contributions:

- Modeling of SRAM cell stability under spatial and temporal V_{th} variations induced by process variation and NBTI stress;
- Dynamic characterization of NBTI induced V_{th} degradation by monitoring the standby leakage current I_{DDQ} ;
- Mitigation methods able to compensate the effects of process variations and NBTI by taking advantage of the FinFET's special device structure.

Simulation results have demonstrated that the proposed technique can effectively maintain the stability of an SRAM array within the desired range during its operational life under both spatial and temporal variations, hence substantially improve the performance and reliability of the system.

Accurate modeling and assessment of system reliability is critical to DRM systems, thus a quantitative reliability model, taking full advantage of the deployed on-chip aging sensors is at premium. Addressing this issue, in Chapter 6 we presented a lifetime reliability modeling and enhancement framework, and the main contributions we have made can be summarized as follows:

- We developed a circuit level delay shift model, which makes use of the collected information from dedicated aging sensors.
- We proposed a Timing-Sharing Sensing (TSS) method for existing V_{th} -based DRM systems meant to combat NBTI-induced wearout in ICs. The TSS method can accurately sample the activity ratio from the circuits under monitoring with a reduced number of aging sensors, which is crucial for the practical implementation of any DRM system.
- We introduce n -Time-To-Failure (n TTF) based "virtual age" concept and utilize it as the system reliability metric, instead of the conventional Mean-Time-To-Failure (MTTF). Compared with MTTF, the virtual age metric sets a more reasonable degradation budget for the system, thus reduces the area and power consumption overhead.

SPICE simulation results have indicated that the proposed circuit reliability model and the TSS method can achieve accurate reliability assessment with

significantly reduced area and power overhead, when compared with existing equivalent state of the art approaches.

The main work of this dissertation covers a wide range of research area, from physical fundamentals of failure mechanisms to high level circuit and architecture designs for reliability management. The central idea behind all the work in this dissertation is to bring a quantitative reliability management system into practice, in order to bridge up the gap between the pessimistic reliability outlook of current and future CMOS technologies and the practical requirements. As the feature size of advanced technology is approaching the physical limitation and devices become more and more fragile to spatial and temporal parameter variations, such kind of reliability management systems will draw more and more attention.

7.2 Future Research Directions

Although we have introduced the general framework of a quantitative reliability-aware computing platform able to perform dynamic reliability management based on aging information collected from in-chip sensors, its practical implementation still requires the further clarifications of many theoretical and practical aspects. In the following, we list several important research directions to further complete and improve the current reliability aware computing platform.

- **Direction 1 - The further understanding of physical fundamentals of failure mechanisms and the compact degradation model for reliability simulation and assessment.**

As we have mentioned in Chapter 2, there are still a lot of arguments on the physical fundamentals on the failure mechanisms, especially for the NBTI degradation. None of the existing degradation models can capture all the NBTI degradation associated phenomenon very well. As a result, the NBTI degradation might be either underestimated or overestimated by the prediction made by current models. More important, most of the current parameter degradation models rely on technology related fitting parameter. The effect of technology scaling on these fitting parameters is not clearly understood. Hence, a parameter calibration is required for every technology modification. Future work is required to reveal the technology dependence of failure mechanisms acceleration factors. Those dependences should be integrated into compact circuit models

(e.g., SPICE models), for reliability simulation at design time and for reliability assessment at runtime.

- **Direction 2 - The effective identification of the most aging-prone critical path at design-time and runtime.**

Regardless the design-time reliability optimization, or the reliability management strategy at the runtime, the effective identification of the most aging-prone critical paths is essential for both cases. For design-time simulations, the device-level reliability model is very time-consuming due to the extreme large population of transistor in designs. For runtime reliability management by monitoring degradation in critical paths, the number of required aging sensors is critical as it may result in substantial area and power overhead. In either case, millions or even billions of possibilities have to be considered. This is further aggravated by the fact that the aging degradation highly depends on the workload in the circuit. In order to perform a more realistic selection of the aging critical path, application-oriented aging profiling should be performed. In this line of reasoning one has to first select a set of representative application benchmarks to simulate the average activity in the circuits. Based on this analysis, aging critical paths should be selected under the stress of activities generated by the application benchmarks.

- **Direction 3 - Statistical quantitative reliability management at runtime.**

If the most aging-prone critical paths can not be effectively identified, an interesting alternative direction could be to select a representative set of logic paths as the circuit-under-monitoring. Instead of monitoring the degradation evolution of circuits in the tail of failure distribution, we monitor the degradation evolution of the failure distribution itself. In such a way, a statistical quantitative reliability management can be achieved according to the failure distribution of the representative set of circuit paths. However, this methodology requires the availability of dynamic degradation distribution information. This kind of information can be obtained from the maintenance information, however, this is not accessible to the academia under the most circumstances.

Bibliography

- [1] *Failure Mechanisms and Models for Semiconductor Devices (JEP122-E)*. JEDEC Publication, 2009.
- [2] BSIM Compact Model for Common(C) Multi-Gate(MG) FETs. <http://www-device.eecs.berkeley.edu/bsim/?page=BSIMCMG>, 2012.
- [3] International Technology Roadmap for Semiconductors. <http://www.itrs.net/>, 2012.
- [4] NanGate 45nm Open Cell Library. <url=http://www.nangate.com/>, 2012.
- [5] Predictive Technology Model (PTM). <http://ptm.asu.edu/>, 2012.
- [6] Arrhenius Equation. url=http://en.wikipedia.org/wiki/Arrhenius_equation, 2013.
- [7] ABADEER, W., AND ELLIS, W. Behavior of NBTI under AC Dynamic Circuit Conditions. In *Proceedings of IEEE International Reliability Physics Symposium (IPRS)* (2003), pp. 17–22.
- [8] ABELLA, J., VERA, X., AND GONZALEZ, A. Penelope: The NBTI-Aware Processor. In *IEEE/ACM International Symposium on Microarchitecture, MICRO* (2007), pp. 85–96.
- [9] ABU-RAHMA, M., AND ANIS, M. *Nanometer Variation-Tolerant SRAM Statistical Design for Yield*. SpringerLink : Bücher. Springer, 2012.
- [10] AGARWAL, M., BALAKRISHNAN, V., BHUYAN, A., AND ET.AL. Optimized Circuit Failure Prediction for Aging: Practicality and Promise. In *IEEE International Test Conference, ITC* (2008), pp. 1–10.
- [11] AGARWAL, M., PAUL, B., ZHANG, M., AND MITRA, S. Circuit Failure Prediction and Its Application to Transistor Aging. In *IEEE VLSI Test Symposium, VTS* (2007), pp. 277–286.
- [12] AGOSTINELLI, M., HICKS, J., XU, J., WOOLERY, B., MISTRY, K., ZHANG, K., JACOBS, S., JOPLING, J., YANG, W., LEE, B., RAZ, T., MEHALEL, M., KOLAR, P., WANG, Y., SANDFORD, J., PIVIN, D., PETERSON, C., DiBATTISTA, M., PAE, S., JONES, M., JOHNSON, S., AND SUBRAMANIAN, G. Erratic Fluctuations of SRAM Cache Vmin at the 90nm Process Technology Node. In *IEEE International Electron Devices Meeting (IEDM) Technical Digest* (2005), pp. 655–658.
- [13] ALAM, M. A Critical Examination of the Mechanics of Dynamic NBTI for PMOS-FETs. In *Proceedings of IEEE International Electron Devices Meeting, IEDM* (2003), pp. 14.4.1–14.4.4.
- [14] ALAM, M., KUFLUOGLU, H., VARGHESE, D., AND MAHAPATRA, S. A comprehensive model for PMOS NBTI degradation: Recent progress. *Microelectronics Reliability* 47, 6 (2007), 853 – 862.
- [15] AMERASEKERA, E. A., AND NAJM, F. N. *Failure Mechanisms in Semiconductor Devices*. J. Wiley, 1997.
- [16] BHARDWAJ, S., WANG, W., VATTIKONDA, R., CAO, Y., AND VRUDHULA, S. Predictive Modeling of the NBTI Effect for Reliable Design. In *Custom Integrated Circuits Conference, CICC* (2006), pp. 189–192.

- [17] BHATTACHARYYA, A. *Compact MOSFET Models for VLSI Design*. Wiley-IEEE Press, 2009.
- [18] BILD, D., BOK, G., AND DICK, R. Minimization of NBTI Performance Degradation Using Internal Node Control. In *Proceedings of Design, Automation Test in Europe Conference Exhibition (2009)*, pp. 148–153.
- [19] BLOME, J. A., FENG, S., GUPTA, S., AND MAHLKE, S. A. Self-Calibrating Online Wearout Detection. In *40th Annual IEEE/ACM International Symposium on Microarchitecture, MICRO (2007)*, pp. 109–122.
- [20] BORKAR, S. Y. Designing Reliable Systems from Unreliable Components: The Challenges of Transistor Variability and Degradation. *Micro Magazine, IEEE* 25, 6 (2005), 10–16.
- [21] C. TOUMAZOU, J.B. HUGHES, AND N.C. BATTERSBY. *Switched-Currents an Analogue Technique for Digital Technology*. Peter Peregrinus Ltd., 1993.
- [22] CHEN, I. C., HOLLAND, S., AND HUT, C. A Quantitative Physical Model for Time-Dependent Breakdown in SiO₂. In *international Reliability Physics Symposium, IRPS (1985)*, pp. 24–31.
- [23] CHOI, Y.-K., HA, D., SNOW, E., BOKOR, J., AND KING, T.-J. Reliability Study of CMOS FinFETs. In *Proceedings of IEEE International Electron Devices Meeting Technical Digest, IEDM (2003)*, pp. 7.6.1 – 7.6.4.
- [24] CHOI, Y.-K., LINDERT, N., XUAN, P., TANG, S., HA, D., ANDERSON, E., KING, T.-J., BOKOR, J., AND HU, C. Sub-20 nm CMOS FinFET Technologies. In *International Electron Devices Meeting Technical Digest, IEDM (2001)*, pp. 19.1.1 –19.1.4.
- [25] CHUNG, J., KO, P.-K., AND HU, C. A Model for Hot-Electron-Induced MOSFET Linear-Current Degradation Based on Mobility Reduction Due to Interface-State Generation. *IEEE Transactions on Electron Devices* 38, 6 (1991), 1362–1370.
- [26] DADGOUR, H., AND BANERJEE, K. Aging-Resilient Design of Pipelined Architectures Using Novel Detection and Correction Circuits. In *Proceedings of Design, Automation Test in Europe Conference Exhibition, DATE (2010)*, pp. 244–249.
- [27] DAS, S., ROBERTS, D., LEE, S., PANT, S., BLAAUW, D., AUSTIN, T., FLAUTNER, K., AND MUDGE, T. A Self-Tuning DVS Processor Using Delay-Error Detection and Correction. *IEEE Journal of Solid-State Circuits* 41, 4 (2006), 792–804.
- [28] DEGRAEVE, R., AOULAICHE, M., KACZER, B., ROUSSEL, P., KAURAUUF, T., SAHHAF, S., AND GROESENEKEN, G. Review of Reliability Issues in High-k/Metal Gate Stacks. In *International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA (2008)*, pp. 1–6.
- [29] DENNARD, R., GAENSSLEN, F., RIDEOUT, V., BASSOUS, E., AND LEBLANC, A. Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions. *IEEE Journal of Solid-State Circuits* 9, 5 (1974), 256–268.
- [30] E. SACKINGER, AND W. GUGGENBUHL. A High-Swing, High-Impedance MOS Cascade Circuit. In *IEEE Journal of Solid-State Circuits (1990)*, pp. 289–298.
- [31] ERSHOV, M., SAXENA, S., KARBASI, H., WINTERS, S., AND ET.AL. Dynamic Recovery of Negative Bias Temperature Instability in P-Type Metal-Oxide-Semiconductor Field-Effect Transistors. *Applied Physics Letters* 83 (2003), 1647.
- [32] FINKELSTEIN, M. *Failure Rate Modelling for Reliability and Risk*. Springer, 2008.

- [33] G. KIM, M.K KIM, B.S. CHANG, AND W. KIM. A Low-Voltage, Low-Power CMOS Delay Element. In *IEEE Journal of Solid-State Circuits* (1996), pp. 966–971.
- [34] GIELEN, G., DE WIT, P., MARICAU, E., AND ETC. Emerging Yield and Reliability Challenges in Nanometer CMOS Technologies. In *Design, Automation and Test in Europe, DATE* (2008), pp. 1322–1327.
- [35] G.O. DUCOUDRAY, R. GONZALEZ-CARVAJAL, AND J. RAMIREZ-ANGULO. A High-Speed Dynamic Current Sensor for IDD test Based on the Flipped Voltage Follower. In *Southwest Symposium on Mixed-Signal Design* (2003), pp. 208–211.
- [36] GRASSER, T., KACZER, B., GOES, W., REISINGER, H., AICHINGER, T., HEHENBERGER, P., WAGNER, P. J., SCHANOVSKY, F., FRANCO, J., LUQUE, M., AND NELHIEBEL, M. The Paradigm Shift in Understanding the Bias Temperature Instability: From Reaction-Diffusion to Switching Oxide Traps. *IEEE Transactions on Electron Devices* 58, 11 (2011), 3652–3666.
- [37] GRASSER, T., WAGNER, P., REISINGER, H., AICHINGER, T., POBEGEN, G., NELHIEBEL, M., AND KACZER, B. Analytic Modeling of the Bias Temperature Instability Using Capture/Emission Time Maps. In *IEEE International Electron Devices Meeting (IEDM)* (2011), IEEE, pp. 27–4.
- [38] GROESENEKEN, G., CRUPI, F., SHICKOVA, A., THIJIS, S., LINTEN, D., KACZER, B., COLLAERT, N., AND JURCZAK, M. Reliability Issues in MuGFET Nanodevices. In *Proceedings of IEEE International Reliability Physics Symposium, IRPS* (2008), pp. 52–60.
- [39] HOUSSA, M., AOULAICHE, M., GENDT, S. D., GROESENEKEN, G., HEYNS, M. M., AND STESMANS, A. Reaction-Dispersive Proton Transport Model for Negative Bias Temperature Instabilities. *Applied Physics Letters* 86, 9 (2005), 093506.
- [40] HU, C., TAM, S. C., HSU, F.-C., KO, P.-K., CHAN, T.-Y., AND TERRILL, K. Hot-Electron-Induced MOSFET Degradation: Model, Monitor, and Improvement. *IEEE Transactions on Electron Devices* 32, 2 (1985), 375 – 385.
- [41] HUANG, L., AND XU, Q. Energy-Efficient Task Allocation and Scheduling for Multi-Mode MPSoCs under Lifetime Reliability Constraint. In *Proceedings of Design, Automation Test in Europe Conference Exhibition, DATE* (2010), pp. 1584–1589.
- [42] HUANG, L., YUAN, F., AND XU, Q. Lifetime Reliability-Aware Task Allocation and Scheduling for MPSoC Platforms. In *Proceedings of Design, Automation and Test in Europe, DATE* (2009), IEEE, pp. 51–56.
- [43] HUANG, X., LEE, W.-C., KUO, C., HISAMOTO, D., CHANG, L., KEDZIERSKI, J., ANDERSON, E., TAKEUCHI, H., CHOI, Y.-K., ASANO, K., SUBRAMANIAN, V., KING, T.-J., BOKOR, J., AND HU, C. Sub 50-nm FinFET: PMOS. In *International Electron Devices Meeting (IEDM) Technical Digest* (1999), pp. 67–70.
- [44] J. KEANE, X. WANG, D. PERSAUD, AND C. KIM. An All-In-One Silicon Odometer for Separately Monitoring HCI, BTI, and TDDB. *IEEE Journal of Solid-State Circuits* 45, 4 (2010), 817–829.
- [45] JEPPESON, K. O., AND SVENSSON, C. M. Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices. *Journal of Applied Physics* 48, 5 (1977), 2004–2014.
- [46] KANG, K., ALAM, M., AND ROY, K. Characterization of NBTI Induced Temporal Performance Degradation in Nano-scale SRAM Array Using IDDQ. In *IEEE International Test Conference, ITC* (2007), pp. 1–10.

- [47] KANG, K., KIM, K., ISLAM, A. E., ALAM, M. A., AND ROY, K. Characterization and Estimation of Circuit Reliability Degradation under NBTI Using On-Line IDDQ Measurement. In *Proceedings of Design Automation Conference, DAC* (2007), pp. 358–363.
- [48] KARL, E., SINGH, P., BLAAUW, D., AND SYLVESTER, D. Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation. In *IEEE International Solid-State Circuits Conference, ISSCC* (2008), pp. 410–623.
- [49] KHAN, O., AND KUNDU, S. A Self-Adaptive System Architecture to Address Transistor Aging. In *Proceedings of Design, Automation Test in Europe Conference Exhibition, DATE* (2009), pp. 81–86.
- [50] KIM, K., AND FOSSUM, J. Double-Gate CMOS: Symmetrical- versus Asymmetrical-Gate Devices. *IEEE Transactions on Electron Devices* 48, 2 (2001), 294–299.
- [51] KIM, K. K., WANG, W., AND CHOI, K. On-Chip Aging Sensor Circuits for Reliable Nanometer MOSFET Digital Circuits. *IEEE Transactions on Circuits and Systems II: Express Briefs* 57, 10 (2010), 798–802.
- [52] KIM, S.-Y., SU PARK, T., LEE, J.-S., PARK, D., KIM, K.-N., AND LEE, J.-H. Negative Bias Temperature Instability (NBTI) of Bulk FinFETs. In *Proceedings of 43rd IEEE International Reliability Physics Symposium, IRPS* (2005), pp. 538–540.
- [53] KIM, T.-H., PERSAUD, R., AND KIM, C. H. Silicon Odometer: An On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circuits. In *IEEE Symposium on VLSI Circuits* (2007), pp. 122–123.
- [54] KUFLUOGLU, H., AND ALAM, M. Theory of Interface-Trap-Induced NBTI Degradation for Reduced Cross Section MOSFETs. *IEEE Transaction on Electron Devices* 53, 5 (2006), 1120–1130.
- [55] KUFLUOGLU, H., AND ASHRAFUL ALAM, M. A Geometrical Unification of the Theories of NBTI and HCI Time-exponents and its Implications for Ultra-scaled Planar and Surround-Gate MOSFETs. In *Proceedings of IEEE International Electron Devices Meeting, IEDM* (2004), pp. 113–116.
- [56] KUMAR, J., BUTLER, K., KIM, H., AND VASUDEVAN, S. Early Prediction of NBTI Effects Using RTL Source Code Analysis. In *ACM/EDAC/IEEE Design Automation Conference, DAC* (2012), pp. 808–813.
- [57] KUMAR, S., KIM, C., AND SAPATNEKAR, S. A Finite-Oxide Thickness-Based Analytical Model for Negative Bias Temperature Instability. *IEEE Transactions on Device and Materials Reliability* 9, 4 (2009), 537–556.
- [58] KUMAR, S., KIM, K., AND SAPATNEKAR, S. Impact of NBTI on SRAM Read Stability and Design for Reliability. In *International Symposium on Quality Electronic Design, ISQED* (2006), pp. 6 pp.–218.
- [59] KUMAR, S. V., KIM, C. H., AND SAPATNEKAR, S. S. An Analytical Model for Negative Bias Temperature Instability. In *Proceedings of IEEE/ACM International Conference on Computer-Aided Design, ICCAD* (New York, NY, USA, 2006), ICCAD '06, ACM, pp. 493–496.
- [60] KUO, K., CHIEN, W., AND KIM, T. *Reliability, Yield, and Stress Burn-In: A Unified Approach for Microelectronics Systems Manufacturing & Software Development*. Kluwer Academic, 1998.

- [61] LACHENAL, D., MONSIEUR, F., REY-TAURIAC, Y., AND BRAVAIX, A. HCI Degradation Model based on the Diffusion Equation including the MVHR Model. *Microelectronics Engineering* 84 (2007), 1921–1924.
- [62] LEE, H., LEE, C.-H., PARK, D., AND CHOI, Y.-K. A Study of Negative-Bias Temperature Instability of SOI and Body-Tied FinFETs. *IEEE Electron Device Letters* 26, 5 (2005), 326 – 328.
- [63] LI, X., QIN, J., AND BERNSTEIN, J. Compact Modeling of MOSFET Wearout Mechanisms for Circuit-Reliability Simulation. *IEEE Transaction on Device and Materials Reliability* 8, 1 (2008), 98–121.
- [64] LIAO, W.-S., LIAW, Y.-G., TANG, M.-C., CHAKRABORTY, S., AND LIU, C. W. Investigation of Reliability Characteristics in NMOS and PMOS FinFETs. *IEEE Electron Device Letters* 29, 7 (2008), 788 –790.
- [65] LORENZ, D., GEORGAKOS, G., AND SCHLICHTMANN, U. Aging Analysis of Circuit Timing Considering NBTI and HCI. In *Proceedings of IEEE International On-Line Testing Symposium, IOLTS* (2009), pp. 3–8.
- [66] LU, D. *Compact Models for Future Generation CMOS*. PhD thesis, EECS Department, University of California, Berkeley, 2011.
- [67] LU, Z., LACH, J., STAN, M., AND SKADRON, K. Improved Thermal Management with Reliability Banking. *IEEE Micro Magazine* 25, 6 (2005), 40–49.
- [68] MAHAPATRA, S., ISLAM, A. E., DEORA, S., MAHETA, V. D., JOSHI, K., JAIN, A., AND ALAM, M. A. A Critical Re-evaluation of the Usefulness of RD Framework in Predicting NBTI Stress and Recovery. In *International Reliability Physics Symposium, IRPS* (2011), pp. 614–623.
- [69] MARICAU, E., AND GIELEN, G. *Analog IC Reliability in Nanometer CMOS*. Springer, 2013.
- [70] MCPHERSON, J. Reliability Trends with Advanced CMOS Scaling and The Implications for Design. In *IEEE Custom Integrated Circuits Conference, CICC* (2007), pp. 405–412.
- [71] MCPHERSON, J., AND BAGLEE, D. Acceleration Factors for Thin Gate Oxide Stressing. In *International Reliability Physics Symposium, IRPS* (1985), pp. 1–5.
- [72] MCPHERSON, J. W. Reliability Challenges for 45nm and Beyond. In *Proceedings of Design Automation Conference, DAC* (2006), DAC '06, pp. 176–181.
- [73] MCPHERSON, J. W. *Reliability Physics and Engineering: Time-To-Failure Modeling*. Springer, 2010.
- [74] MINTARNO, E., SKAF, J., ZHENG, R., VELAMALA, J., CAO, Y., BOYD, S., DUTTON, R., AND MITRA, S. Optimized Self-Tuning for Circuit Aging. In *Proceedings of Design, Automation Test in Europe Conference Exhibition, DATE* (2010), pp. 586–591.
- [75] MOENS, P., VARGHESE, D., AND ALAM, M. Towards a Universal Model for Hot Carrier Degradation in DMOS Transistors. In *International Symposium on Power Semiconductor Devices IC's (ISPSD)* (2010), pp. 61–64.
- [76] MONTGOMERY, D., AND RUNGER, G. *Applied Statistics and Probability for Engineers*. John Wiley & Sons, 2010.
- [77] MOORE, G. E. Cramming More Components onto Integrated Circuits. *Electronics* 38, 8 (Apr. 1965), 114–117.

- [78] NDAI, P., BHUNIA, S., AGARWAL, A., AND ROY, K. Within-Die Variation-Aware Scheduling in Superscalar Processors for Improved Throughput. *IEEE Transactions on Computers* 57, 7 (2008), 940–951.
- [79] OGAWA, S., AND SHIONO, N. Generalized diffusion-reaction model for the low-field charge-buildup instability at the si-sio₂ interface. *Physical Review B* 51 (1995), 4218–4230.
- [80] OHRING, M. *Reliability and Failure of Electronic Materials and Devices*. Elsevier Science, 1998.
- [81] O’UCHI, S., MASAHARA, M., SAKAMOTO, K., ENDO, K., LIU, Y., MATSUKAWA, T., SEKIGAWA, T., KOIKE, H., AND SUZUKI, E. Flex-Pass-Gate SRAM Design for Static Noise Margin Enhancement Using FinFET-Based Technology. In *IEEE Custom Integrated Circuits Conference, CICC* (2007), pp. 33–36.
- [82] PARK, J., PARK, J.-M., SOHN, S.-O., LEE, J.-B., JEON, C.-H., HAN, S. Y., YAMADA, S., YANG, W., ROLL, Y., AND PARK, D. Reliability Investigations for Bulk-FinFETs Implementing Partially-Insulating Layer. In *Proceedings of 45th Annual IEEE International Reliability Physics Symposium, IPRS* (2007), pp. 378 –381.
- [83] PARTHASARATHY, C., DENAIS, M., HUARD, V., RIBES, G., VINCENT, E., AND BRAVAIX, A. Characterization and Modeling NBTI for Design-in Reliability. In *IEEE International Integrated Reliability Workshop Final Report* (2005), pp. 5 pp.–.
- [84] PARTHASARATHY, C., DENAIS, M., HUARD, V., RIBES, G., VINCENT, E., AND BRAVAIX, A. New Insights into Recovery Characteristics Post NBTI Stress. In *Proceedings of IEEE International Reliability Physics Symposium* (2006), pp. 471–477.
- [85] PARTHASARATHY, C., DENAIS, M., HUARD, V., RIBES, G., VINCENT, E., AND BRAVAIX, A. New Insights Into Recovery Characteristics During PMOS NBTI and CHC Degradation. *IEEE Transactions on Device and Materials Reliability* 7, 1 (2007), 130–137.
- [86] PAUL, B., KANG, K., KUFLUOGLU, H., ALAM, M., AND ROY, K. Impact of NBTI on the Temporal Performance Degradation of Digital Circuits. *IEEE Electron Device Letters* 26, 8 (2005), 560–562.
- [87] PHAM, H. *Springer Handbook of Engineering Statistics*. Springer, 2006.
- [88] POBEGEN, G., AICHINGER, T., NELHIEBEL, M., AND GRASSER, T. Understanding temperature acceleration for NBTI. In *IEEE International Electron Devices Meeting (IEDM)* (2011), IEEE, pp. 27–3.
- [89] RAMACHANDRAN, P., ADVE, S. V., BOSE, P., AND RIVERS, J. A. Metrics for Architecture-Level Lifetime Reliability Analysis. In *Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS* (2008), IEEE, pp. 202–212.
- [90] RAUCH, S.E., I. The Statistics of NBTI-Induced V_t and β Mismatch Shifts in pMOS-FETs. *IEEE Transactions on Device and Materials Reliability* 2, 4 (2002), 89–93.
- [91] RAUCH, S. E., AND LA ROSA, G. The Energy Driven Paradigm of nMOSFET Hot Carrier Effects. In *Proceedings of 43rd Annual IEEE International Reliability Physics Symposium, IRPS* (2005), IEEE, pp. 708–709.
- [92] RISCH, L. Pushing CMOS Beyond the Roadmap. *IEEE Journal of Solid-State Electronics* 50, 4 (2006), 527 – 535. Papers Selected from the 35th European Solid-State Device Research Conference - ESSDERC’05.

- [93] SAKURAI, T., AND NEWTON, A. Alpha-Power Law MOSFET Model and Its Applications to CMOS Inverter Delay and Other Formulas. *IEEE Journal of Solid-State Circuits* 25, 2 (1990), 584–594.
- [94] SHAH, N., SAMANTA, R., ZHANG, M., HU, J., AND WALKER, D. Built-In Proactive Tuning System for Circuit Aging Resilience. In *IEEE International Symposium on Defect and Fault Tolerance of VLSI Systems, DFTVS* (2008), pp. 96–104.
- [95] SHANG, H., CHANG, L., WANG, X., ROOKS, M., ZHANG, Y., TO, B., BABICH, K., TOTIR, G., SUN, Y., KIEWRA, E., IEONG, M., AND HAENSCH, W. Investigation of FinFET Devices for 32nm Technologies and Beyond. In *Proceedings of International Symposium on VLSI Technology Digest* (2006), pp. 54–55.
- [96] SINGH, P., KARL, E., SYLVESTER, D., AND BLAAUW, D. Dynamic NBTI Management using a 45nm Multi-Degradation Sensor. In *IEEE Custom Integrated Circuits Conference, CICC* (2010), pp. 1–4.
- [97] SRINIVASAN, J., ADVE, S., BOSE, P., AND RIVERS, J. The Case for Lifetime Reliability-Aware Microprocessors. In *International Symposium on Computer Architecture, ISCA* (2004), pp. 276–287.
- [98] SRINIVASAN, J., ADVE, S. V., BOSE, P., AND RIVERS, J. A. Exploiting Structural Duplication for Lifetime Reliability Enhancement. In *International Symposium on Computer Architecture, ISCA* (2005), IEEE Computer Society, pp. 520–531.
- [99] STOLK, P., WIDDERSHOVEN, F., AND KLAASSEN, D. Modeling Statistical Dopant Fluctuations in MOS Transistors. *IEEE Transactions on Electron Devices* 45, 9 (1998), 1960–1971.
- [100] STRONG, A., WU, E., VOLLERTSEN, R., SUNE, J., ROSA, G., SULLIVAN, T., AND STEWART E. RAUCH, I. *Reliability Wearout Mechanisms in Advanced CMOS Technologies*. IEEE Press Series on Microelectronic Systems. Wiley, 2009.
- [101] SUN, J., KODI, A., LOURI, A., AND WANG, J. NBTI Aware Workload Balancing in Multi-Core Systems. In *Proceedings of Quality of Electronic Design, ISQED* (2009), pp. 833–838.
- [102] SUN, J., LYSECKY, R., SHANKAR, K., KODI, A., LOURI, A., AND WANG, J. Workload Capacity Considering NBTI Degradation in Multi-Core Systems. In *Proceedings of Asia and South Pacific Design Automation Conference, ASP-DAC* (2010), pp. 450–455.
- [103] SYLVESTER, D., BLAAUW, D., AND KARL, E. ElastIC: An Adaptive Self-Healing Architecture for Unpredictable Silicon. *IEEE Design & Test of Computers* 23, 6 (2006), 484–490.
- [104] TAKEDA, E., AND SUZUKI, N. An Empirical Model for Device Degradation due to Hot-Carrier Injection. *IEEE Electron Device Letters* 4, 4 (1983), 111–113.
- [105] TAM, S., KO, P.-K., AND HU, C. Lucky-Electron Model of Channel Hot-Electron Injection in MOSFET'S. *IEEE Transactions on Electron Devices* 31, 9 (1984), 1116–1125.
- [106] TENBROEK, B., LEE, M., REDMAN-WHITE, W., BUNYAN, J., AND UREN, M. Self-heating Effects in SOI MOSFETs and their Measurement by Small Signal Conductance Techniques. *IEEE Transactions on Electron Devices* 43, 12 (1996), 2240–2248.
- [107] TEODORESCU, R., AND TORRELLAS, J. Variation-Aware Application Scheduling and Power Management for Chip Multiprocessors. In *International Symposium on Computer Architecture, ISCA* (2008), pp. 363–374.

- [108] TISDALE, W. A., WILLIAMS, K. J., TIMP, B. A., NORRIS, D. J., AYDIL, E. S., AND ZHU, X.-Y. Hot-Electron Transfer from Semiconductor Nanocrystals. *Science* 328, 5985 (2010), 1543–1547.
- [109] TIWARI, A., SARANGI, S. R., AND TORRELLAS, J. ReCycle: Pipeline Adaptation to Tolerate Process Variation. In *International Symposium on Computer Architecture, ISCA* (2007), pp. 323–334.
- [110] TIWARI, A., AND TORRELLAS, J. Facelift: Hiding and Slowing Down Aging in Multicores. In *Proceedings of the International Symposium on Microarchitecture (MICRO)* (2008).
- [111] V. REDDY, A.T. KRISHNAN, A. MARSHALL, ET.AL. Impact of Negative Bias Temperature Instability on Digital Circuit Reliability. In *Proceedings of International Reliability Physics Symposium* (2002), pp. 248 – 254.
- [112] WANG, W., REDDY, V., KRISHNAN, A. T., VATTIKONDA, R., KRISHNAN, S., AND CAO, Y. Compact Modeling and Simulation of Circuit Reliability for 65-nm CMOS Technology. *IEEE Transaction on Device and Materials Reliability* 7, 4 (2007), 509–517.
- [113] WANG, W., REDDY, V., YANG, B., BALAKRISHNAN, V., KRISHNAN, S., AND CAO, Y. Statistical Prediction of Circuit Aging under Process Variations. In *IEEE Custom Integrated Circuits Conference, CICC* (2008), pp. 13–16.
- [114] WANG, W., WEI, Z., YANG, S., AND CAO, Y. An Efficient Method to Identify Critical Gates under Circuit Aging. In *Proceedings of IEEE/ACM International Conference on Computer-Aided Design, ICCAD* (2007), pp. 735 –740.
- [115] WANG, Y., CHEN, X., WANG, W., CAO, Y., XIE, Y., AND YANG, H. Leakage Power and Circuit Aging Cooptimization by Gate Replacement Techniques. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 19, 4 (2011), 615–628.
- [116] WANG, Y., WANG, Y., TARR, G., AND INIEWSKI, K. A Temperature, Supply Voltage Compensated Floating-Gate MOS Dosimeter Using VTH Extractor. In *Proceedings of International Workshop on System-on-Chip for Real-Time Applications* (2005), pp. 176 – 179.
- [117] WU, E., AITKEN, J., NOWAK, E., VAYSHENKER, A., VAREKAMP, P., HUECKEL, G., MCKENNA, J., HARMON, D., HAN, L.-K., MONTROSE, C., AND DUFRESNE, R. Voltage-Dependent Voltage-Acceleration of Oxide Breakdown for Ultra-Thin Oxides. In *Technical Digest International Electron Devices Meeting, IEDM* (2000), pp. 541–544.
- [118] WU, E., HARMON, D., AND HAN, L.-K. Interrelationship of Voltage and Temperature Dependence of Oxide Breakdown for Ultrathin Oxides. *IEEE Electron Device Letters* 21, 7 (2000), 362–364.
- [119] ZHUO, C., SYLVESTER, D., AND BLAAUW, D. Process Variation and Temperature-Aware Reliability Management. In *Design, Automation Test in Europe Conference Exhibition, DATE* (2010), DATE '10, pp. 580 –585.

List of Publications

International Journals

1. **Y. Wang**, M. Enachescu, S. D. Cotofana, L. Fang, **Variation Tolerant On-Chip Degradation Sensors for Dynamic Reliability Management Systems**, *Microelectronics Reliability*, 2012(52), pp. 1787–1791.
2. **Y. Wang**, S. D. Cotofana, L. Fang, **Analysis of the Impact of Spatial and Temporal Variation on the Stability of SRAM Arrays and Mitigation Technique Using Independent-Gate Devices**, accepted by *Journal of Parallel and Distributed Computing* (in press).

International Conference Proceedings

1. N. Cucu Laurenciu, **Y. Wang**, S. D. Cotofana, **A Direct Measurement Scheme of Amalgamated Aging Effects with Novel On-Chip Sensor**, to appear in the proceedings of IPF/IEEE International Conference on Very Large Scale Integration (VLSI-SoC), Istanbul, Turkey, October, 2013.
2. **Y. Wang**, S. D. Cotofana, L. Fang, **Lifetime Reliability Assessment with Aging Information from Low-Level Sensors**, to appear in *Proceedings of Great Lakes Symposium on VLSI*, Paris, France, May 2013.
3. **Y. Wang**, S. D. Cotofana, L. Fang, **Statistical Reliability Analysis of NBTI Impact on FinFET SRAMs and Mitigation Technique Using Independent-Gate Devices**, *Proceedings of 2012 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pp. 109–115, Amsterdam, The Netherlands, July 2012.
4. **Y. Wang**, S. D. Cotofana, L. Fang, **A Unified Aging Model of NBTI and HCI Degradation towards Lifetime Reliability Management for Nanoscale MOSFET Circuits**, *Proceedings of IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*, pp. 175–180, San Diego, United States of America, 2011.
5. **Y. Wang**, S. D. Cotofana, L. Fang, **A Novel Virtual Age Reliability Model for Time-to-Failure Prediction**, *IEEE International Integrated Reliability Workshop Final Report (IIRW)*, pp. 109–115, South Lake Tahoe, United States of America, 2010.

Samenvatting

Agressieve CMOS technologie feature size schaling is al gaande sinds afgelopen decennia, terwijl de voedingsspanning niet proportioneel mee schaal. Door de toenemende vermogensdichtheid en elektrische veld in de gate dielektricum, zijn versnellende factoren van faalmechanismen in nanoschaal Integrated Circuits (ICs) ernstiger dan ooit geworden. Als gevolg daarvan, wordt het handhaven van IC betrouwbaarheid op gewenste niveau een kritische uitdaging tijdens zowel design-time en runtime. Dit proefschrift onderzoekt betrouwbaarheid-bewust ontwerp en beheertechnieken om de betrouwbaarheid en kwaliteit van de IC-producten te garanderen. Met onze speciale belangen voor de tijdsafhankelijke device parameter degradaties als gevolg van intrinsieke faalmechanismen, richten we onze discussie op: (i) beoordeling van runtime betrouwbaarheid, (ii) aging degradaties, en (iii) mitigatie technieken die betrouwbaarheid-bewuste berekeningen mogelijk te maken. Daartoe stellen we een dynamisch Reliability Management (DRM) framework voor om de genduceerde aging degradatie tegen te gaan. Om een kwantitatief beheer te realiseren, zijn toegewijde online aging sensoren tewerk gesteld in het voorgestelde framework om dynamisch degradatie informatie uit circuits af te leiden. We stellen eerst een uniform aging model voor, voor de opkomende FinFET devices als de fysieke basis voor het begrijpen van de onderliggende aging degradatie. Vervolgens introduceren we twee soorten van aging sensoren gebaseerd op threshold spanning en stroommeting van de stroomvoorziening, respectievelijk, om te helpen bij de online betrouwbaarheidsbeoordeling van DRM-systemen. Vervolgens introduceren we een compensatie techniek om de 6T SRAM cel stabiliteit te beheren onder ruimtelijke en temporele variaties, door threshold voltage modulatie met behulp van back-gate biasing van onafhankelijke-gate FinFET devices. We concluderen het proefschrift met het presenteren van een levenslange betrouwbaarheid model en enhancement framework, dat laat zien hoe aging informatie gebruikt kan worden komend van speciaal low-level aging sensoren om de algemene IC gezondheidstoestand binnen de voorgeschreven grenzen te houden.

Curriculum Vitae



Yao WANG was born on July 21st, 1983 in Xiangtan, Hunan Province, China. From 2001 to 2005 he studied at the School of Computer Science and Technology, National University of Defense Technology (NUDT) in Changsha, China. He received Bachelor's degree in Computer Science and Technology in 2005 and his Master's degree in Electronic Science and Technology in 2007, from National University of Defense Technology.

In 2008 he was awarded a government scholarship from Chinese Scholarship Council (CSC), to pursue his PhD studies in the Netherlands. In October 2008, he joined the Computer Engineering laboratory of Delft University of Technology in the Netherlands, under the supervision of Associate Professor Dr. Sorin Cotofana. The major focus of his PhD studies are on reliability-aware resource allocation and management towards nanoscale computation. The results of this work are presented in the current dissertation.

Yao's research interests include physical modeling of failure mechanisms in nanoscale device, dynamic reliability management on multicore and many-core platforms, novel nanoscale electronic devices, and architecture for future computing, fault-tolerant, and reliability-critical applications.