Statistical Characterization and Decomposition of SRAM cell Variability and Aging

by

Venkatesa Ravi

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved February 2013 by the
Graduate Supervisory Committee:

Yu Cao, Chair
Bertan Bakkaloglu
Lawrence Clark

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

Memories play an integral role in today's advanced ICs. Technology scaling has enabled high density designs at the price paid for impact due to variability and reliability. It is imperative to have accurate methods to measure and extract the variability in the SRAM cell to produce accurate reliability projections for future technologies. This work presents a novel test measurement and extraction technique which is non-invasive to the actual operation of the SRAM memory array. The salient features of this work include i) A single ended SRAM test structure with no disturbance to SRAM operations ii) a convenient test procedure that only requires quasi-static control of external voltages iii) non-iterative method that extracts the VTH variation of each transistor from eight independent switch point measurements.

With the present day technology scaling, in addition to the variability with the process, there is also the impact of other aging mechanisms which become dominant. The various aging mechanisms like Negative Bias Temperature Instability (NBTI), Channel Hot Carrier (CHC) and Time Dependent Dielectric Breakdown (TDDB) are critical in the present day nano-scale technology nodes. In this work, we focus on the impact of NBTI due to aging in the SRAM cell and have used Trapping/De-Trapping theory based log(t) model to explain the shift in threshold voltage VTH. The aging section focuses on the following i) Impact of Statistical aging in PMOS device due to NBTI dominates the temporal shift of SRAM cell ii) Besides static variations , shifting in VTH demands increased guard-banding margins in design stage iii) Aging statistics remain constant during the shift, presenting a secondary effect in aging prediction. iv) We have investigated to see if the aging mechanism can be used as a compensation technique to

reduce mismatch due to process variations. Finally, the entire test setup has been tested in SPICE and also validated with silicon and the results are presented. The method also facilitates the study of design metrics such as static, read and write noise margins and also the data retention voltage and thus help designers to improve the cell stability of SRAM.

DEDICATION

To my family and friends

ACKNOWLEDGMENTS

I like to start by thanking Dr. Yu Cao for having given me the opportunity to be a part of the research community and for also giving me a chance to work in NIMO group. The numerous weekly meetings gave the enriching experience of doing research and my advisor's valuable guidance and support has helped me learn the discipline towards research. Through the two years spent in the lab, I have had the opportunity to work with different people whose contributions have played significant role in shaping up my thesis. It my humble pleasure to convey my gratitude to all of them.

I am grateful to Dr. Lawrence Clark and Dr. Bertan Bakkaloglu for agreeing to be a part of my committee and also for taking time to review my work.

My sincere thanks to Ketul Sutaria, Jyothi Bhaskarr Velamala and Naveen Suda for their valuable feedback and encouragement while at NIMO group. I also thank them for their contributions towards this research work. It was a pleasure working with them all.

Last but not the least, all this wouldn't have been possible if not for my family and my friends who have been a constant source of motivation in helping me realizing this Master's thesis.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER 1

INTRODUCTION

1.1 History of Semiconductors.

First transistor and subsequently the integrated circuit certainly qualify as two great inventions of the twentieth century. The accident invention of the first point contact transistor in 1947 by John Bardeen, Walter Brittain and William Shockley revolutionized semiconductor research and paved way to the new era in semiconductor industry replacing the bulky and inefficient vacuum tubes [1]. This sparked the new age of modern technical accomplishments from space explorations to portable electronic devices today. Packing these transistors compactly lead to the development of first ever integrated circuit by Jack Kilby in 1958 at Texas Instruments. This lead to the steady progress in the semiconductor

Figure1.1. Evolution of modern day ICs from vacuum tubes.

industry with Fairchild releasing its first ever commercial IC in 1963 consisting of two logic gates called 907. This was later followed by the Micromosaic consisting of a few hundred devices resembling a FGPA or PAL of the modern day [1]. These transistors were to be programmed by software to implement any connectivity and function. Figure1.1 shows the evolution of transistors starting from vacuum tubes to modern day ICs.

1.2 Technology Scaling

Quoting the famous Moore's Law "The number of transistors on a chip doubled every 18 to 24 months" [2]. This indeed has been the trend since it was first stated back in 1965 in the article "Cramming more components onto integrated circuits". Contrary to it being a prediction, it slowly emerged to become truth over decades due to the rapid scaling of technology. The predictive statement was responsible for the rapid shrink in device sizes over the decades leading to the present day ultra-deep submicron technology. Figure 1.2 shows the trend predicted by Gordon Moore. It clearly emphasizes the fact that, the number of transistors packed into an integrated chip has indeed increased tremendously in the past decades.

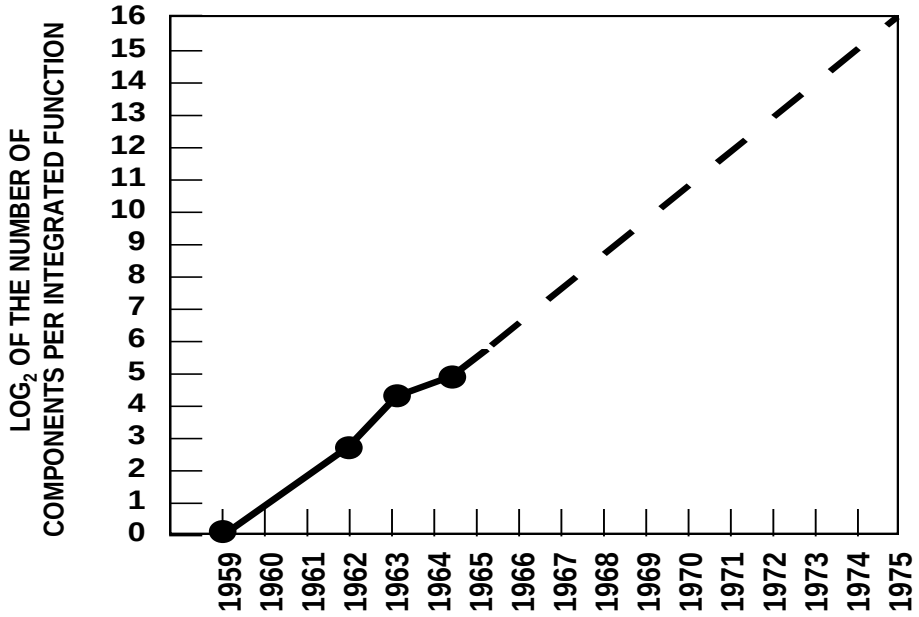

Figure1.2. Transistor count vs years with technology scaling as predicted by Moore's Law. Courtesy: Electronics, Volume 38, Number 8, April 19, 1965

The need for more portable electronic devices and gadgets has driven the need for technology scaling enabling fabrication of smaller geometries into a single wafer with improved performance and power exacerbating the effects of noise and reliability issues. Delving

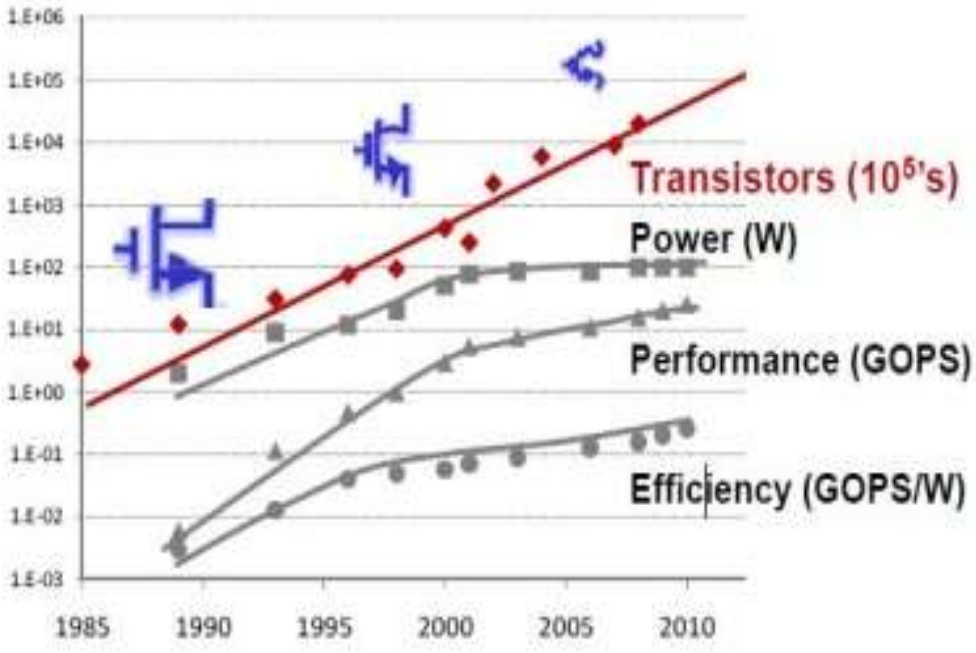

Figure1.3. Impact on performance metrics due to technology scaling.
Courtesy: [T. Mudge, U. Mich]

deeper down to the sub-micron nodes has brought to light the disasters of high power consumption, leakage and various forms of reliability concerns which will be discussed in the next section. It is necessary to mitigate these down falls of technology scaling keeping with the benefits from the same [3]. It is evident that there has been grand success in scaling from the following perspectives, i) Increased no. of transistors in a single chip, ii) Increase operation frequency iii) reduced cost. However, the pitfall associated with these successes is the price paid for the reduced reliability and yield of the circuits, poor battery life, and high power dissipation impacting the design for the future. Figure 1.3 illustrates the updated

picture due to scaling [4]. The technology scaling shrinks by 0.7 for every generation approximately. With every new generation, we have the capability of integrating 2x more transistors with marginal increase in the chip fabrication cost. The cost of function decreases by 2x for every generation. The question of how to embed more functionality into a chip is answered by more efficient design methods by exploring the different level or hierarchies of abstraction from transistor to system level.

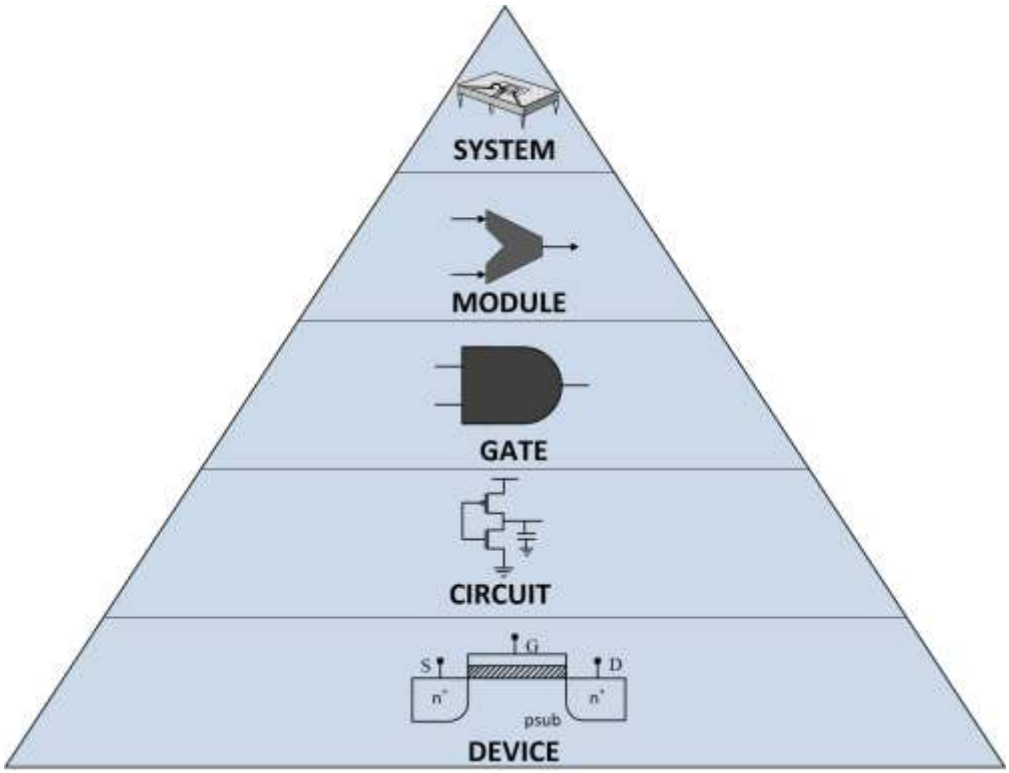

Figure 1.4 Different Design Abstraction levels.

1.3 Reliability and Variability: Impact of Aging

The rapid evolution of Silicon process technologies presents significant challenges to the understanding of the physics of failure of circuits and the characterization of their reliability. It is becoming a daunting challenge to improve reliability with each successive technology generation since the time to failure due to most of the degradation physics decreases. Complicating the problem is the introduction of newer materials and device structures, increasing the risk that either the existing or new failure mechanisms will constrain the pace of scaling. The forefront of the problem is to mitigate these reliability issues by improving the design of manufacturing process. The two popular approaches are i) development of detailed failure physics by understanding the microscopic perspective of the failure mechanism eliminating the defects involved in optimizing the manufacturing process and develop close to accurate compact models for the reliability projections, ii) establishing a linkage between the various degrading 'unit elements', i.e. transistors, interconnects and degradation of circuit performance. This paves way to study the degradation mechanism at unit level and leverage it at the circuit design level.

The continuous drive for technology scaling to integrate more devices into a single chip for new age portable electronic has exacerbated the subject of failures of circuits. This also invites various impacts and effects due to reliability and variability caused by the scaling. Various effects due to variability are defined below concluding with figure 1.5 showing the various components.

a) Layout dependent stress [5]: Increase in doping concentration to due to scaling to match high drive current in turn is impacted due to high scattering of large number of dopant atoms present.

Layout dependent mechanical stress techniques introduced to mitigate the mobility degradation due to increase dopant concentration.

b) High-K gate effects (HK/MG) [6]: Replacing SiO_2 as replacement to gate dielectric is most urgent for low power circuits where gate leakage related power consumption is serious limitation. The charges trap in the interfacial layer. The fixed charge within the high-k film shifts V_{TH}, relative to that for SO2, and poses a serious issue for V_{TH}, control that must be solved for production of circuits using high-k gate dielectric.

c) Random dopant fluctuations (RDF) [7]: Due to process variation from implanted impurity concentration, the random dopant fluctuations can cause significant impact on the transistor characteristics like altering the threshold voltage. In low sub-micron devices the total number of dopants is fewer and the addition or deletion of even a few impurity atoms cause significant changes. It is a local to each device and hence two juxtaposed devices can have altogether different characteristics due to RDF.

d) Line edge roughness (LER) [8]: Due to the limitations in lithography process, the printed gates exhibit a certain roughness. This is called line edge roughness and can cause fluctuations in transistor parameters which cause degraded yield and performance of the device.

e) Random telegraph noise (RTN): Scaling Technology throws light on RDF and superimposed RTN. RTN occurs on top of other non-idealities, where RDF is one of the most relevant. RTN induces fluctuations on electrical behavior which may change from one instant in time to other. These transient electrical behaviors may induce jitter in oscillators or transient failures in SRAMs. The interplay between RDF and RTN may exacerbate its impact on circuit reliability.

These take effect post fabrication and can be categorized as fresh impacts on the circuit reliability due to technology scaling. In addition to these fresh impacts, there is also aging that comes into picture with the operation time. This manifests itself by altering the device characteristics and subsequently leading to circuit failure and degradation.

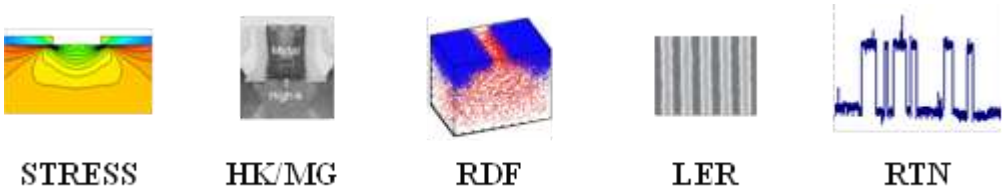

Figure1.5. Variability effects observed due to technology scaling.

Circuit performance deteriorates over time due to aging effects like Negative Bias Temperature Instability (NBTI), Channel Hot Carrier (CHC), and Time Dependent Dielectric Breakdown (TDDB). These effects have taken a main track in the present day nano-scale regime due to changes in the manufacturing process and the scaling, introduced to improve device lifetime and circuit performance. These aging effects manifest itself as change in the device properties like threshold voltage, drain current and so on leading to circuit performance degradation and ultimately design failure. With increased scaling, these aging effects become more severe and it is required to take steps to mitigate the same to reap the benefits of technology scaling. When subject to operating bias, transistors exhibit changes in transistor characteristics over time an effect termed as Bias Temperature Instability (BTI). Typically, transistor thresholds (V_{TH}) increase, and other electrical parameters, such as drive current (I_D) and trans-conductance (Gm), are also affected. One of the major concerns is the NBTI, which causes a shift in the threshold voltage of p-channel transistor with time. It is caused by the

positive charge build-up at the interface between the gate dielectric and channel. Figure 1.6 shows diagram of a typical PMOS NBTI degradation and recovery process well reported in the literature [9]. The fundamental difficulty lies with the increasing variability of NBTI failure times as geometries reduce [10].

Paradigm shift in assessing the reliability of digital circuits is made feasible due to the knowledge in the interaction between the process variability and NBTI, directing us to a higher abstraction i.e. circuit centric approach from the traditional transistor based approach. With the improvements in the manufacturing process to account for the technology scaling, the present day advanced ICs are designed to take process variations induced timing delay into account. As a result, it is only important to consider the circuits that fall out of the initial process distribution as a result of NBTI to present reliability risk [12].

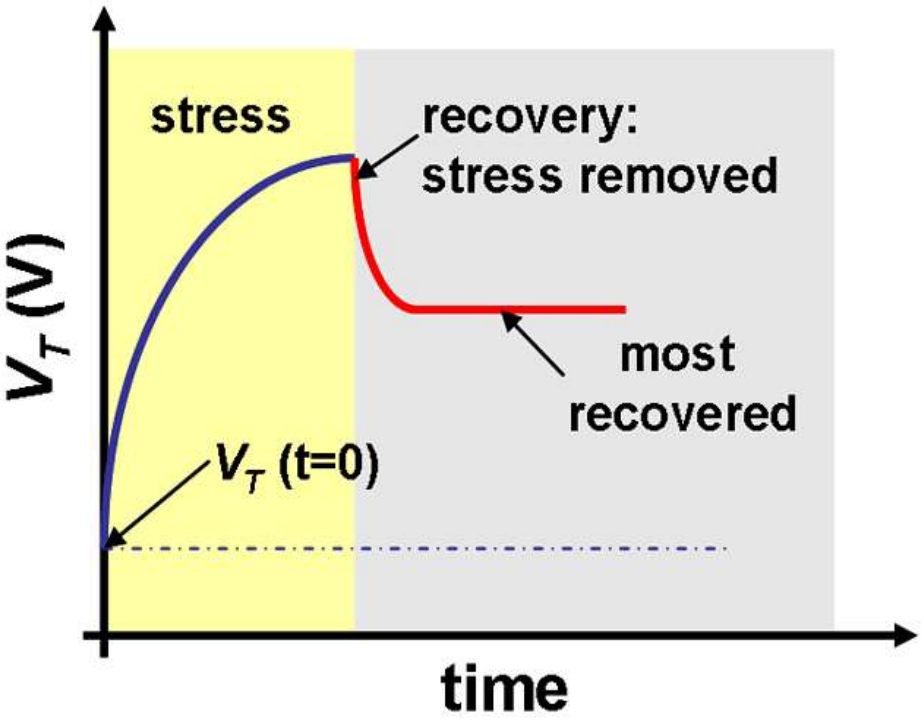

Figure1.6. PMOS NBTI vs Time illustrating both degradation and recovery. Courtesy: Intel

1.4 Role of Memory and Impact due to Reliability.

6-Transistor SRAM cells are ubiquitous in modern day ICs. They comprise over 86% of the transistor count in the server class microprocessors [12]. This significant fraction of chip area occupied by SRAM cells makes the cell area a major indicator of manufacturing capability that provides a key competitive advantage. Due to their use of the smallest geometries possible, SRAMs are especially prone to random microscopic variations such as random dopant fluctuations (RDF), line edge roughness (LER) etc. Aggressive scaling of CMOS technology also brings forth multiple reliability issues such as Negative Bias Temperature Instability (NBTI, Channel Hot Carrier (CHC) along with issues of variability due to process.

These effects manifest as increase in the threshold voltage (V_{TH}) of PMOS and NMOS devices respectively as the circuit lifetime progresses [13]. The sensitivity to these atomistic fluctuations further increases when low power supply voltage is used to reduce active and leakage power. Therefore, the design of SRAM cell under process variability, voltage and temperature is a tremendous challenge that is overcome by extensive experimentation in test silicon. The various performance metrics like Static Noise Margin (SNM) is very sensitive to device mismatch which arises due to the process variability as a result of technology scaling. The need for more memory in smaller area in the present day portable devices like cellphones, music players, and tablets has led to very aggressive scaling of the SRAM memory arrays and increased the sensitivity to aging related effects like NBTI induced transistor V_{TH} mismatch in addition to already present process mismatch. The bit cell has traditionally reduced 2x every two years as shown in the Figure 1.7, which translates to pointing that the bit counts are also increasing at a

corresponding rate. As a result various Error Correction Codes (ECC) or techniques are adopted to improve the Vmin (V_{DD}) characteristics [9].

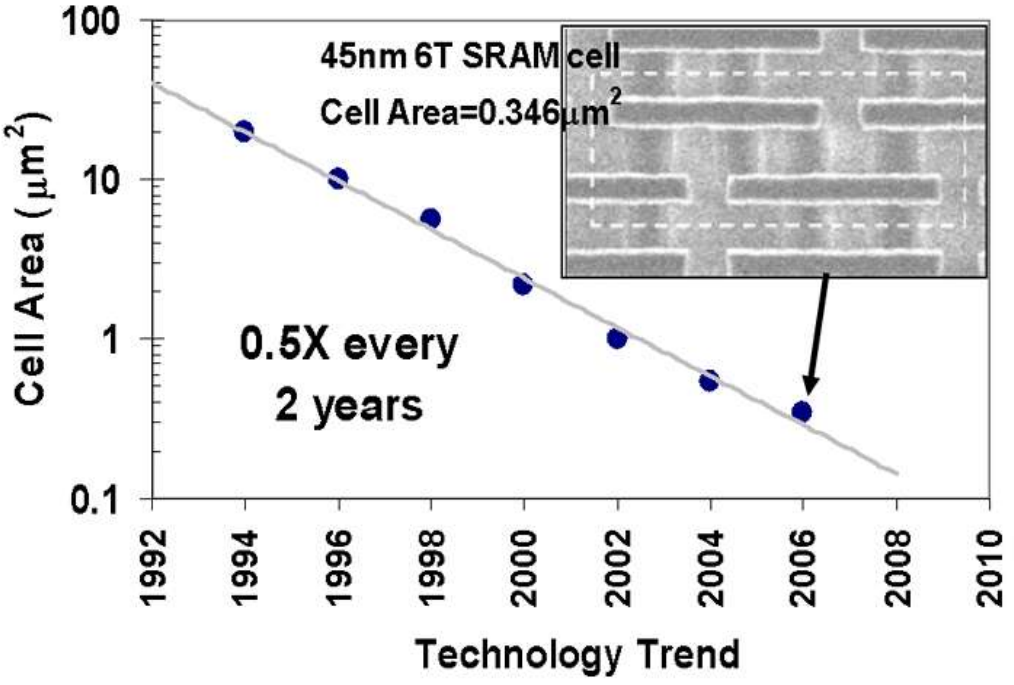

Figure1.7. 6T SRAM cell size scaling trend showing 2X cell area scaling every two years Courtesy: Intel

1.5 Previous Works

An efficient and reliable method is required to ensure SRAM manufacturability and yield and to determine the magnitude of variations in SRAM cells. The test method should not interfere with the actual operation of the SRAM cell and also be extendable to large volume commercial SRAM arrays. A large number of techniques today help to perform low yield analysis by electrically probing the SRAM cells to characterize their quality for both process development and high volume IC production. Given below are a few approaches tried by researchers earlier and published.

i) An individual cell is chosen and voltage applied to one of the bit lines (BL) node and the voltage on the other BL node (BLN) is swept while the current is measured. [14].

ii) An alternative method of write margin extraction consists of applying a voltage sweep on the word line (WL) from 0 to V_{DD} and monitoring the internal node voltage. [15,16]

These test methods allow electrical characterization of the specific defect types. Specially designed test structures help probe the internal nodes with additional pads with increase in area costs and even affect the regular operation of the cell. Some methods probe the SRAM Static Noise Margin (SNM), the minimum data retention voltage (DRV) or the full N-curves [17]. However, the downside of these approaches is that, it is difficult to extract the actual as fabricated transistor level variations for further diagnosis.

1.6 Thesis contributions

It is well known that at the present day CMOS technology node, the geometries are so small that, they have brought to light the enormous impact from reliability and yield perspective. We are aware that SRAM array use special absolutely small geometries as compared to the logic circuits to enable dense packing of memory in the given chip with minimal impact on overall chip area. Due to this, the influence of scaling has created a bigger challenge to overcome the ill effects of process variability and device aging on memory arrays. The main objective of this thesis is to develop a test methodology to probe the nodal voltages in a SRAM array without affecting its regular operation and use it to decompose and find the variation in the threshold voltage in each device in each cell. This will help predict the performance metrics of the memory array and improve to compensate the reliability effect both due process variation and aging.

In this method, we show that with appropriate SRAM voltage choices, the V_{TH} of each SRAM cell can be extracted by probing the BL responses in a voltage mode, which is less impacted by on chip leakage currents. It is a non-invasive procedure which does not disturb the data stored in the memory during the test measurement. 8 different voltage samples of BL and BLN are measured and used to decompose the variation in the V_{TH} of the 6 devices in the SRAM cell.

To show the impact of aging in addition to the process variation, aged stress measurements is made in commercial 90nm process and the data is used to extract the V_{TH} shift due to aging and also statistical aging analysis is done to predict the SRAM cell margins. The method involved choosing voltage threshold values during measurement to simplify extraction procedure by using simple KCL.

Finally the thesis concludes with the study of the various performance metrics of the SRAM memory array and how they are affected due to the aging and process variations.

1.7 Thesis Organization

Chapter 1 discusses the introduction and motivation behind the entire thesis work on memory reliability and techniques to predict its performance. Chapter 2 elucidates the test measurement setup, the extraction procedure along with the various transistor models used in the technique. Chapter 3 discusses the validation performed to test the robustness of the procedure with spice and 90nm silicon data and results presented. Chapter 4 introduces the impact of aging in SRAM memory array including a small section to show the compact modeling for NBTI. It also includes the results that match with the aged silicon measurement in the 90nm process and presents clearly the shift in V_{TH} due to aging as a result of NBTI. Chapter 5 concludes the thesis with the results and findings and also discusses the scope of future work. It is followed by the reference and appendix showing the matlab codes, solver and spice files used to validate the entire test procedure.

CHAPTER 2

TEST PRINCIPLE AND METHODOLOGY

2.1 Overview of test method

The main idea of this approach is to ensure that the existing SRAM operation is unaffected by the test procedure. In this work an integral test method is proposed and developed to characterize the variability of each cell and efficiently extract the variation of the individual SRAM cell transistors [12]. The salient features of this test procedure are as follows:

i) A cell based test and extraction method that is scalable to any volume of SRAM banks.

ii) Single ended test structure which is non-invasive to the SRAM cell, minimizing disturbance to SRAM topology and function by test.

iii) Quasi-static test which required eight sample points per cell, which is achieved by probing the V_{TH} variations through external voltage control.

iv) Non-iterative method which efficiently decomposes the V_{TH} variation of each transistor from the eight measurements.

v) Easy integration to existing SRAM design and test without need of additional complex test structures.

This entire approach is demonstrated in a 90nm test chip with 32K cells. This procedure also enables accurate prediction of the SRAM performance variability. The extracted V_{TH} are also verified by silicon data for cell write margins and data retention voltage with error less than 3%.

2.2 Test Principle

To probe individual V_{TH} variations in a 6-T SRAM cell, a series of simultaneous write and read operations are conducted under various WL supply voltages. Figure 2.1 illustrated the experimental setup and the preferred bias voltages for a single SRAM cell. The cell itself is biased at a lower supply voltage (V_{DDCELL}), the reason to which will be apparent in the coming sections.

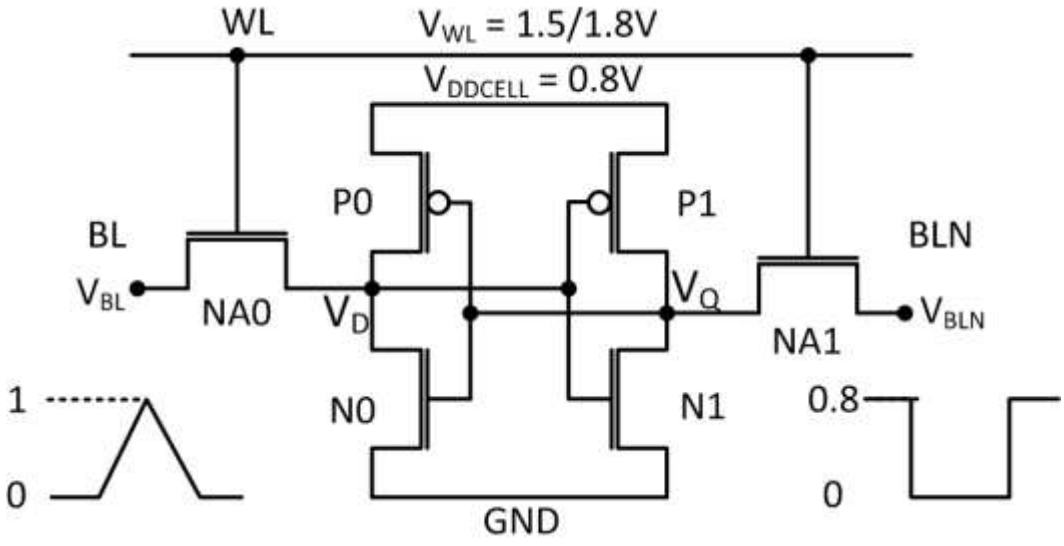

Figure2.1Singled ended test circuit showing different power supply voltages to sample SRAM cell switch points.

Due to inherent nature of the SRAM cell design, write operations are normally performed differentially rather than in single ended fashion due to the inability to write logic 1 through the NMOS access transistor (NA0 and NA1 in Figure2.1). The read stability criteria ensures that the pull down NMOS (N0 and N1) are the stronger than the access devices to prevent accidental flipping of the cell during the read process [3]. In order to identify the voltage required to flip the

state of the cell (both high and low), it is required to do a single ended write of both 1 and 0. This also helps us to measure the ratio of the strengths between the pull up and pull down vs. the access devices. As a result we have to destabilize the cell which is accomplished by using different supply voltages, since we don't want to disturb or interfere with the array structure and the transistor size or layout.

It is known that the order of sizing in a SRAM cell begins with the pull down devices being the strongest followed by the access device and finally the pull up devices are the weakest. To destabilize the cell, it is required to make the access device stronger than the pull down and this is achieved by the higher WL supply

Figure2.2 SRAM test die photo and the overlay of the SRAM bank test structure.

voltage than the cell supply voltage (i.e. $V_{DD} > V_{DDCELL}$). The higher voltage at the gate of the access device helps over the V_{OV} ($V_{GS} - V_{TH}$) loss, thus making it stronger than pull down NMOS and enabling a single ended write. The SRAM cell core is the back to back inverter pair at supply voltage V_{DDCELL} which is made equal to V_{DD} at normal operations. V_{DD} and V_{DDCELL} are independent of each other in the test mode so that the signal strengths are controlled

independently. Figure 2.2 shows the SRAM test die photo and the overlay of the SRAM bank test structure.

2.3 SRAM Cell Design

The test chip used in this work is fabricated in a commercial 90nm CMOS process. It is a 32K cells array allowing direct electrical measurement of the SRAM cell characteristics. Aspect ratios of the transistors (W/L nm) are N0, N1 = 190nm/100nm, P0, P1 = 120nm/100nm and NA0, NA1 = 130nm/130nm. This setup provides full access to the SRAM cell storage nodes. The test is essentially DC in nature without any time dependence. The BL and BLN voltages accurately represent that of the SRAM cell internal nodes Q and QB. For simplicity, the sense amplifier and other peripheral components are omitted in the figure and the explanation. Modification has to be done to perform in-situ measurement of the SRAM transistor characteristics. As a result of the modifications, the test array provides full access to the SRAM cell storage nodes, by separating the power supply of the SRAM cell and the word line (WL). To allow direct forcing and monitoring of the cell switching, multiplexers are added to the bit line (BL). Accurate test equipment and sufficiently small time steps eliminate the time dependencies leaving less chances of error in the measurement.

2.3 Test Procedure

The measurement is performed using two National Instruments data acquisition boards PCI 6229 and PCI 6541. PCI 6229 controls the BL and BLN signals while PCI 6541 provides clock, address and other control signals. The BL and BLN are externally tri stated using discrete analog multiplexer, so they may be both inputs and outputs.

The test proceeds as follows: There are a total of eight switch point measurements for each cell. Two sets of 4 measurements are done at different WL supply voltages. BL and BLN are alternatively driven from 0 to 1V and then from 1 to 0V in steps of 1mV for each cell. When BL/BLN is the input the other bitline BLN/BL serves as the output. The point where the cell state transitions across a specified threshold voltage is recorded. For the case of rising signal from 0 to 1V, a threshold voltage of 0.7 is fixed at the output (BL/BLN). When the output (BL/BLN) transitions from 0.8 to 0V, the value of input (BL/BLN) is noted when the output reaches 0.7V. The same procedure is repeated by alternating the inputs and outputs between BL and BLN. In the case of falling signal the threshold at output is taken as 0.2V.

Hence, for each cell, at a given V_{DD} and V_{DDCELL} pair, four switch points are obtained: two for each rising and falling cases of the input in BL and BLN. However, these four points will only aid in assessing the relative strength ratios of the bit cell Pull up vs. Access and Pull down vs. Access devices. In order to accurately decompose the ΔV_{TH} of each device, more switch points are needed. The entire procedure is repeated for another WL supply voltage is measured thus giving us a total of eight switch points for each cell. These eight independent switch point measurements at WL 1.5V and 1.8V are used to

Switch Case	Input	Output	V_{DD} (V)	Threshold @ Output(V)
SW1	BL rising	BLN	1.8	0.7
SW2	BL falling	BLN	1.8	0.2
SW3	BLN rising	BL	1.8	0.7
SW4	BLN falling	BL	1.8	0.2
SW5	BL rising	BLN	1.5	0.7
SW6	BL falling	BLN	1.5	0.2
SW7	BLN rising	BL	1.5	0.7
SW8	BLN falling	BL	1.5	0.2

Table 2.1 various switch points along with the threshold at outputs.

accurately decompose the six independent ΔV_{TH} in the bit cell. Table 2.1 below shows the various switch case points along with the threshold fixed for each case at the output.

Figure 2.3 shows the PDF for all the eight switch point case measured for the 32K cells in 90nm CMOS process. The results are repeatable under the same condition to within 1-2mV, proving the repeatability of the test apparatus and procedure. We have also reversed BL and BLN in our test apparatus, obtaining essentially identical results.

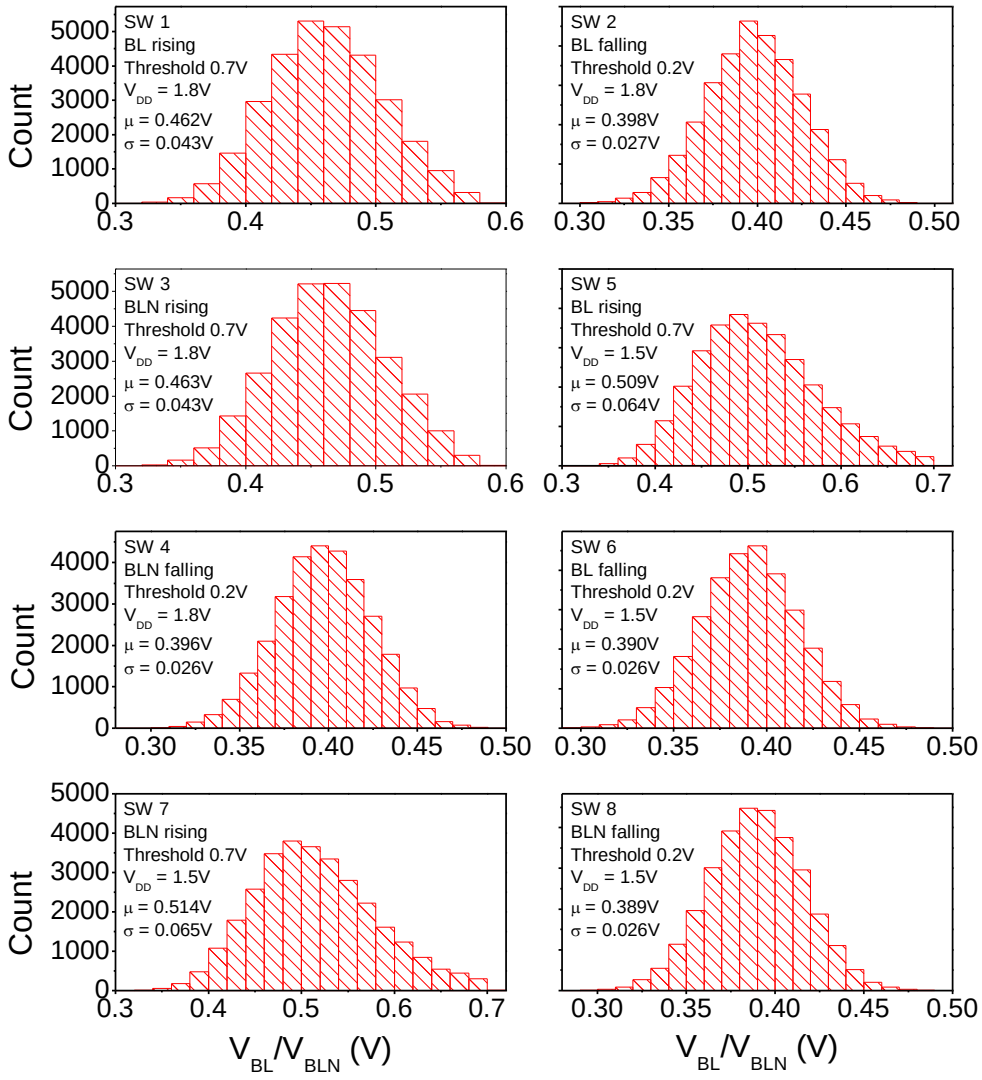

Figure2.3 PDF distribution of Measured Switch points for 32k cells from the commercial 90nm silicon.

2.5 Non-Iterative Extraction Procedure

The eight sample points are used to decompose the ΔV_{TH} of each transistor in the SRAM cell [18]. The extracted V_{TH} is critical for process improvement and design optimization. It also aids process engineers to comprehend the variation characteristics to adaptively minimize them. This also helps in predicting the various performance metrics of a SRAM cells viz. Static Noise margins (Read, Write and Hold) and also the data retention voltage (DRV) thereby enabling design of more robust memory arrays. There are many ways to back trace and extract the V_{TH} variation in each cell. One of the fundamental approaches is using the optimization statement in SPICE to iteratively fit the variations. However, we have six V_{TH} values and eight independent switch point measurements making it an over-determined system making the optimization problem challenging to solve requiring complex numerical methods. In addition the non-linear behavior of the devices worsens the extraction. A standard Linux workstation may take up to 8 hours to solve only 100 cells, with up to 12% of the cases failing to converge.

The method followed in this work brings a novel numerical method to solve and extract the V_{TH} of the 6 devices. The concept of the flow is based on identifying the region of operation of each of the six devices during the read and write operation. The switching thresholds mentioned earlier are chosen such that, we can simplify the number of equations and independent variables involved in the extraction process and formulate the extraction problem to be as linear as possible. This helps in solving the six ΔV_{TH} from right sample points non-iteratively. The time complexity is $O(n)$ where n is the number of cells.

The procedure is based on the physical understanding of the SRAM transistor biases and circuit operating conditions. The method works on one cell at a time independent of the other cells and also doesn't assume any distribution function of V_{TH} variations. There are two different cases as mentioned earlier in this test setup, one is the input rising case and the other is the input falling case. Both the cases are discussed in detail below along with the transistor models.

Case A: Rising Input signal

Let's for simplicity assume we are dealing with BL as the input and BLN as the output for the purpose of elaborating on the technique. During the BL switching, the currents in the cell follow the following KCL.

$$I_{N0} = I_{NA0} + I_{P0} \quad (2.1)$$

$$I_{N1} = I_{NA1} + I_{P1} \quad (2.2)$$

As BL switches, the voltages at node BLN is observed, the voltage at BLN is equal to the voltage in the internal V_Q node since the measurement is essentially DC. Thus, the current through the access device NA1 is 0 ($I_{NA1} = 0$). This reduces the Eq (2.2) to balance only the currents in N1 and P1. On the other hand, the currents in the other side of the cell represented by the Eq (2.1) still hold true and none of the device currents are negligible. The currents in these devices hold across sub-threshold, linear and saturation regions complicating the numerical solution and posing a significant challenge to efficient extraction.

Since our extraction method is based on identifying the operating conditions of the different transistors in the SRAM cell to solve for the V_{TH}, we select appropriate thresholds at the output during the measurement to simplify the Eq. (2.1) in similar lines to Eq. (2.2). If the threshold is chosen close to V_{DDCELL} at the output side, the PMOS device on the left inverter

(since BL is the input side) will be biased in deep sub-threshold region and we will safely be able to neglect the sub-threshold current and thus the Eq. (2.1) (i.e. $I_{NA0} = I_{N0}$) is greatly simplified. In this experiment, we have chosen the threshold to be 0.7V which is very close the cell V_{DD} (0.8V). In this 90nm process it has been observed that the current in deep sub-threshold in I_{P0} is less than 0.1% of I_{N0} which clearly justifies our decision. We can choose any value for the output threshold as long as we ensure that the PMOS device is biased in deep sub-threshold region.

Since, we have made some assumptions the equations presented earlier simplify to the following.

$$I_{N0} = I_{NA0} \tag{2.3}$$

$$I_{N1} = I_{P1} \tag{2.4}$$

It is very evident that I_{NA0} is always biased in strong linear region due to the high word line voltage while I_{N0} is biased either in the saturation or linear region depending on the specific V_{TH} of the N0. The transistor models for NA0 and N0 used to fit these are linear in either case and are presented below in Eq. (2.5 and 2.6) respectively.

$$I_{NA0} = K_{na}[(V_{WL} - \beta V_{BL} - \lambda_n V_{thn})(V_D - V_{BL})] \tag{2.5}$$

$$I_{N0} = K_n[V_D - \eta_n V_{thn} + \alpha_n V_Q] \tag{2.6}$$

Here β, λ_n, η_n, α_n, K_{na}, K_n are fitting coefficients from SPICE simulations. During fitting the V_Q voltage is fixed as 0.7V in our experiment and the V_{WL} is varied between 1.5 and 1.8V to ensure proper fitting across the 8 switch points. Coming to the other inverter devices, they operate in boundary between the sub-threshold and the saturation depending on the individual V_{TH} values. A single exponential function for P1 and N1 in Eq. (2.7 and 2.8) respectively is introduced to smoothly connect these two regions, which is scalable with V_D and V_{TH}.

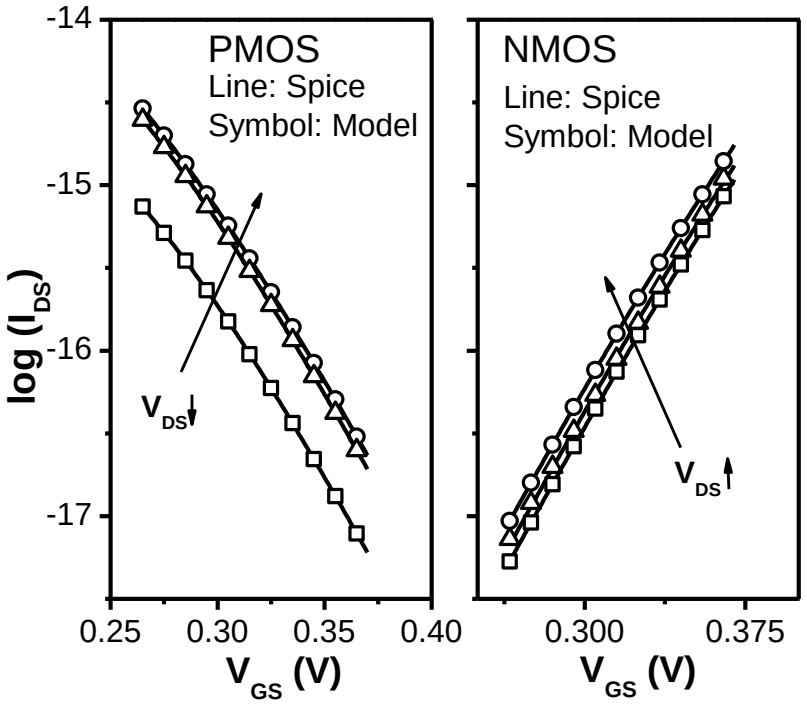

Figure2.4 I_D-V_{GS} fitting within 2% error for N and P devices for different values of V_{DS}

$$I_{P1} = A_p e^{[(V_{DD}-V_D)-\gamma_p V_{thp}+\theta_p (V_{DD}-V_Q)]/d_p} \tag{2.7}$$

$$I_{N1} = A_n e^{[(V_D)-\gamma_n V_{thn}+\theta_n (V_Q)]/d_n} \tag{2.8}$$

A_n, A_p, γ_p, θ_p, d_p, d_n, A_n, γ_n, θ_n, are fitting coefficients from SPICE simulations. Piece-wise model approach is followed for minimal fitting error and the proposed models well match with SPICE simulations with less than 2% error as shown in Figure 2.4.

The advantage of having an exponential function is that, it is easy to linearize it by taking logarithm on both sides. Together with the definition of V_Q at 0.7V, this step effectively converts the KCL Eq. (2.1) into a linear function of V_{TH} and the node voltages. This enables us to solve the problem without iteration.

Case B: Falling Input signal

In this second case, the following KCLs hold true.

$$I_{P0} = I_{NA0} + I_{N0} \quad (2.9)$$

$$I_{N1} = I_{NA1} + I_{P1} \quad (2.10)$$

Going by the same assumption as stated previously the current through I_{NA1} is almost 0 and hence Eq. (2.10) simplifies to the form where the currents in I_{N1} equates to the current in I_{P1}. For Eq. (2.9) we are required to choose threshold voltage at output such that one of the devices of the left inverter is biased into deep sub-threshold region. In this case, we choose the output threshold to be 0.2V there by pushing the NMOS (N0) to operate in deep sub-threshold region.

The PMOS (P0) is modeled with the following current model since it operates in either saturation or linear region depending on the V_{TH} of the device.

$$I_{P0} = K_p[(V_{DD} - V_D) - \eta_p V_{thp} + \alpha_p (V_{DD} - V_Q)] \quad (2.11)$$

The reduced KCL is given by

$$I_{P0} = I_{NA0} \quad (2.12)$$

$$I_{N1} = I_{P1} \quad (2.13)$$

Based on the above equations and models, we can solve for the six independent V_{TH} using the switch point measurements. Since the derivation involves a second order term in V_D due to feedback loop in the SRAM, eight measurements are minimum required to solve without iterating.

2.6 Summary

The flowchart below summarizes the method in solving and decomposing the V_{THs} of all

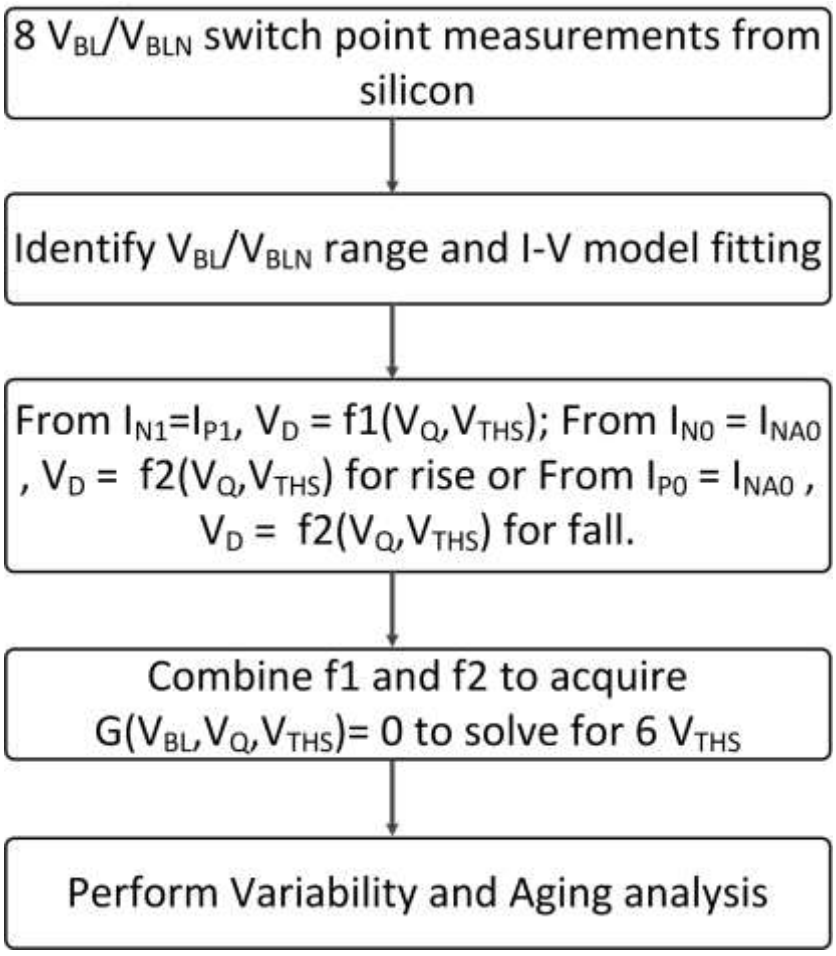

Figure2.5 Flow of steps involved in V_{TH} extraction and aging framework

the devices. Eight Switch point measurements are first ordered to identify the range of nodal voltages to do a piecewise fitting of the transistor models with minimal error. For each of the V_{BL} measurement, we first solve for V_D from I_{N1}=I_{P1} equations and determine V_D as a function of V_Q and V_{TH} (V_D = F1[V_Q,V_{TH}]). In the case of input rising, we solve V_D from the equation I_{NA0} = I_{N0} and find V_D in terms of V_Q and V_{TH} (V_D = F2 [V_Q, V_{TH}]). For the falling input case, we use

the equation $I_{NA0} = I_{P0}$ and find V_D (V_D = F3 [V_Q, V_{TH}]). The above solutions of V_D are used to create a function consisting of only V_{BL}, V_Q, V_{TH}. We know the values of V_{BL} from the measurements and V_Q is a user defined threshold and hence the final framework is reduced to have V_{TH} as the only variable. The first stage of identifying and fitting the models is done using SPICE simulations and Matlab. The fitting is done using first order regression using matrix division so that the fitting is as linear as possible. The final function obtained is solved using a Matlab solver with specified error threshold and value convergence thresholds. It takes about 20 minutes to run on a standard PC to solve and extract V_{TH} for 32K cells.

Recently published work [19] indicates that V_{TH} is a function of operating region and the variation in threshold voltage for sub-threshold region is higher than in strong inversion (saturation/linear) region. This helps us do the extraction to a higher degree of accuracy. Hence, first a coarse extraction is performed and the V_{THs} of all the devices in all the cells are obtained. The mean of the threshold voltages from this extraction is used as input to perform a fine grain extraction which directly gives us the ΔV_{TH} for all the devices in each cell. During the fine extraction the variation in V_{TH} due to operating region is accounted for using some fitting parameters.

In summary, the mapping from V_{BL} to V_{TH} is achieved by solving the current equations at nodes D and Q. The selection of threshold value of V_Q is essential to simplify the derived current equations and enable efficient solutions. This entire procedure is validated in SPICE and also tested in 90nm silicon data as discussed in the following chapter.

CHAPTER 3

VALIDATION

3.1 Verification with SPICE

The extraction methodology provides designers with helpful insight to analyze memory design for variability and aging. Thus, it is very important to validate the threshold voltage extraction method. In this chapter, a comprehensive validation of the threshold voltage extraction method using 90nm silicon data from commercial process and as well as SPICE based approach is presented.

The accuracy of the extraction procedure is first demonstrated by a simulation setup which emulates the silicon measurement setup. Randomly generated V_{TH} variations ($\sigma \sim 60mV$) were injected in 1K the 6-T SRAM cell devices. The pseudo static measurement was emulated using a DC sweep and the V_{BL}/V_{BLN} values were simulated. Based on the simulation values, the I-V models were fitted with piece wise linear approach to have minimal error (< 2%). The extraction procedure is run using these simulated V_{BL}/V_{BLN} values and the ΔV_{TH} are extracted. Correlation between the injected and extracted ΔV_{TH} with RMS error < 3% is shown in Figure 3.1. As expected the correlation is linear between the injected and extracted, thereby establishing the accuracy of the extraction approach. The error in NMOS and PMOS variability extraction is less compared to that of the Access transistor V_{TH} extraction. This proves that, the extraction algorithm followed is able to give precise values of V_{TH} variations and hence can be trusted upon with values extracted from silicon measurements.

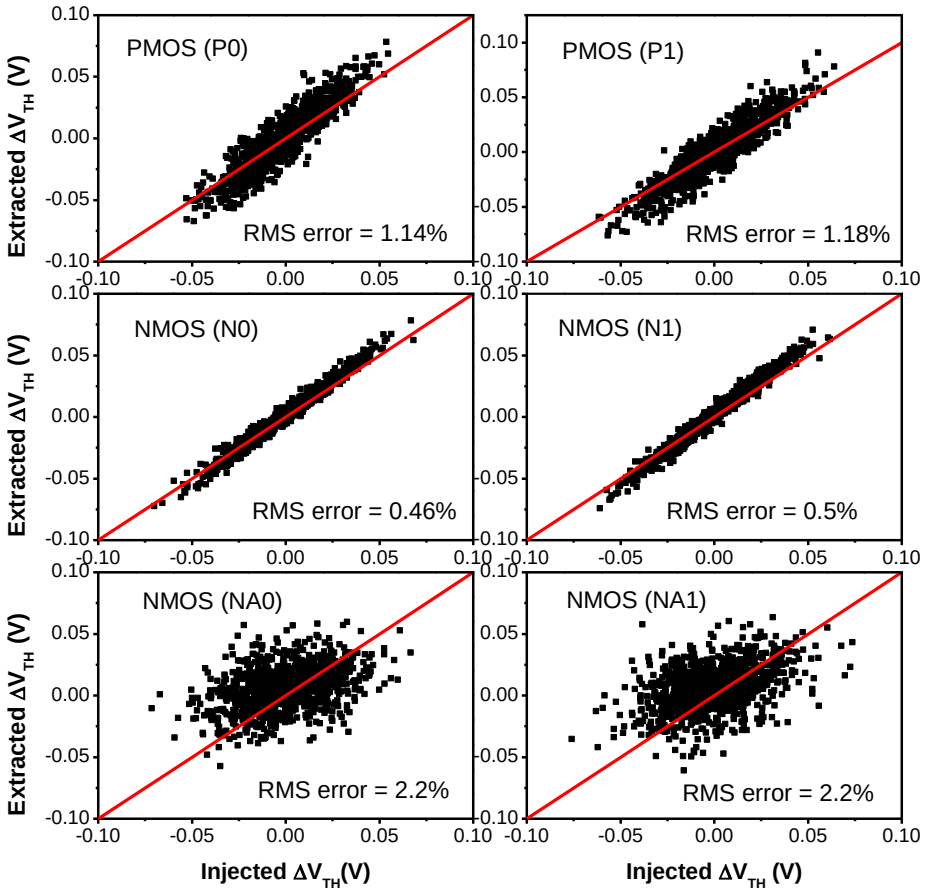

Figure3.1 Correlation between injected V_{TH} in the SPICE environment vs. V_{TH} extracted using extraction methodology for each device in 6T-SRAM cell

Furthermore, to validate the prediction accuracy of performance metrics such as DRV, DRV is simulated with injected V_{TH} and validated with the simulated DRV using extracted V_{TH}. Figure 3.2 shows that the extraction method well predicts the variation in DRV with μ = 103mV and σ = 41mV.

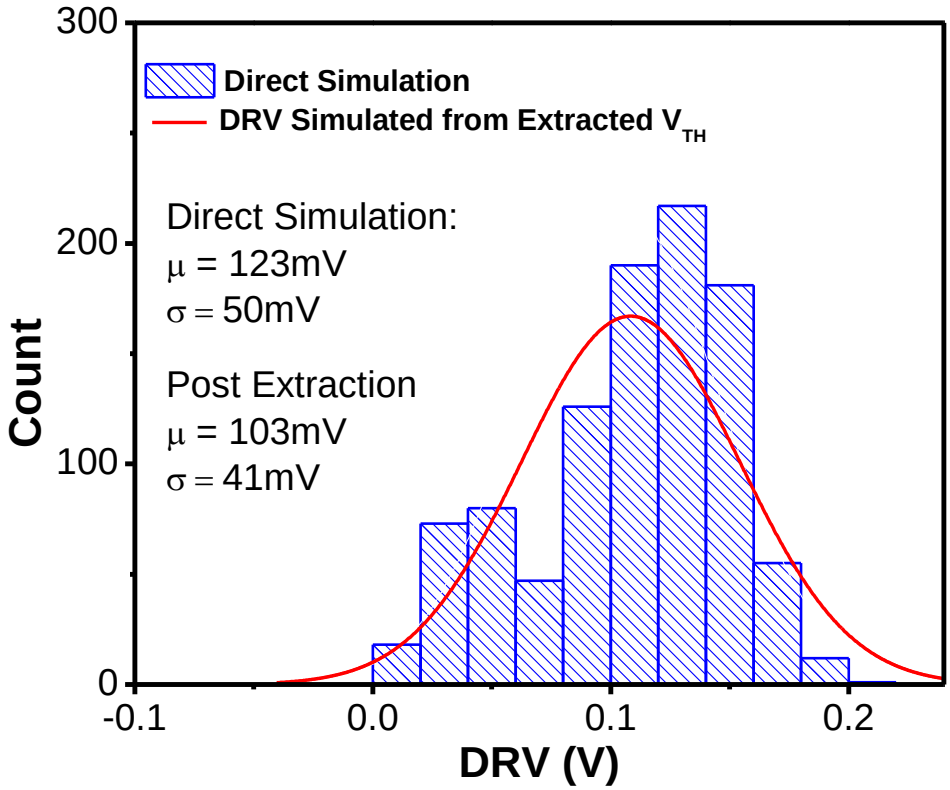

Figure3.2 DRV match for injected V_{TH} values and the extracted V_{TH}. Mean of both DRV matches accurately validating V_{TH} extraction approach

3.2 Silicon Validation

After verifying the procedure with simulation based framework, the switch point measurements from the 90nm silicon data was used to do extract the V_{TH} variations in the 32K cell SRAM array. The original I-V fitting carried out for the simulation framework was calibrated based on the V_{BL}/V_{BLN} values from silicon. Even though the calibration adds error to the system during extraction, it is very negligible and is overcome by the repetitive convergence nature of the solver used for the extraction. Figure 3.3 shows the distribution of the V_{TH} of each device in the SRAM cell for 32K cells. The total time taken to decompose each V_{TH} for such a big array is under 20 minutes using a standard workstation PC. This fast and efficient method is way lesser than the traditional complex numerical methods earlier used.

Most of the distributions are nearly Gaussian, reflecting the fact that V_{TH} variation in a SRAM cell is mainly induced by random variation sources e.g. RDF or LER. The numerical method provided the SRAM cell V_{TH} for each transistor. This extracted V_{TH} enables the designers to predict the trend in the various SRAM performance metrics like Noise margins, data retention voltage and so on.

The performance metric validation is also performed with silicon data. The measured DRV of the 4K cells was verified with SPICE simulation of DRV using the extracted V_{TH} from the extraction procedure. Figure 3.4 shows the correlation of the silicon vs. simulation DRV data for

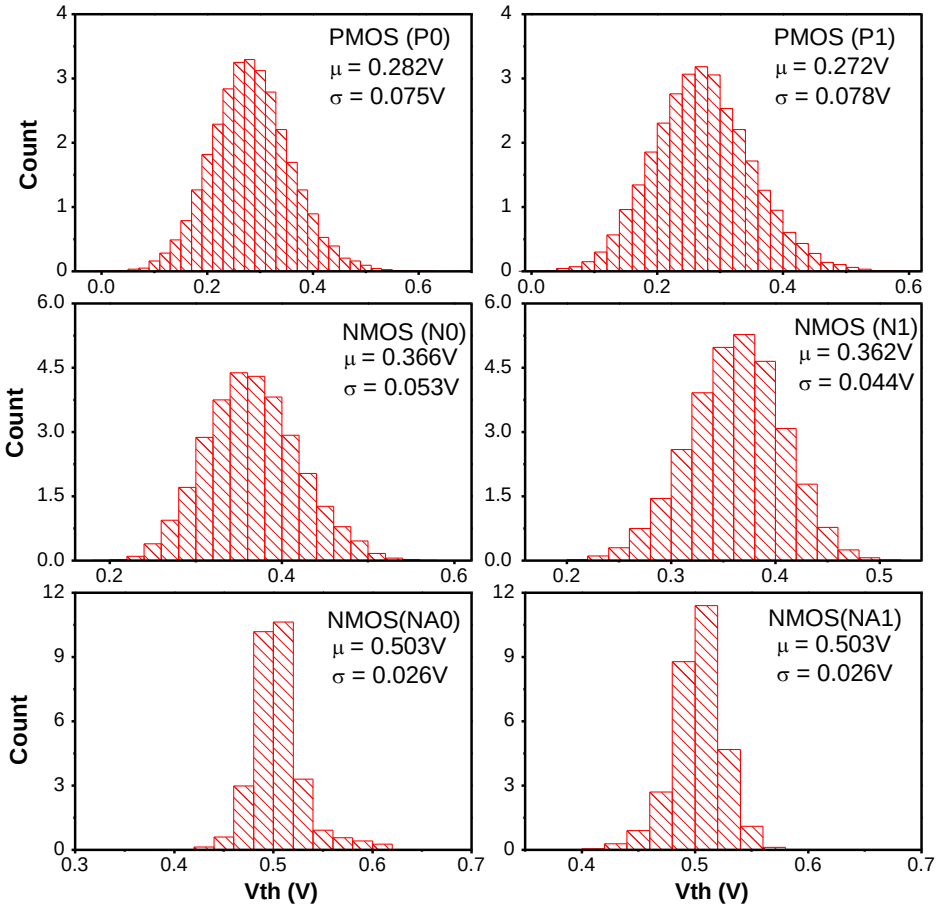

Figure3.3 V_{TH} values extracted using proposed methodology for the 90nm commercial process.

the 90nm SRAM test chip.

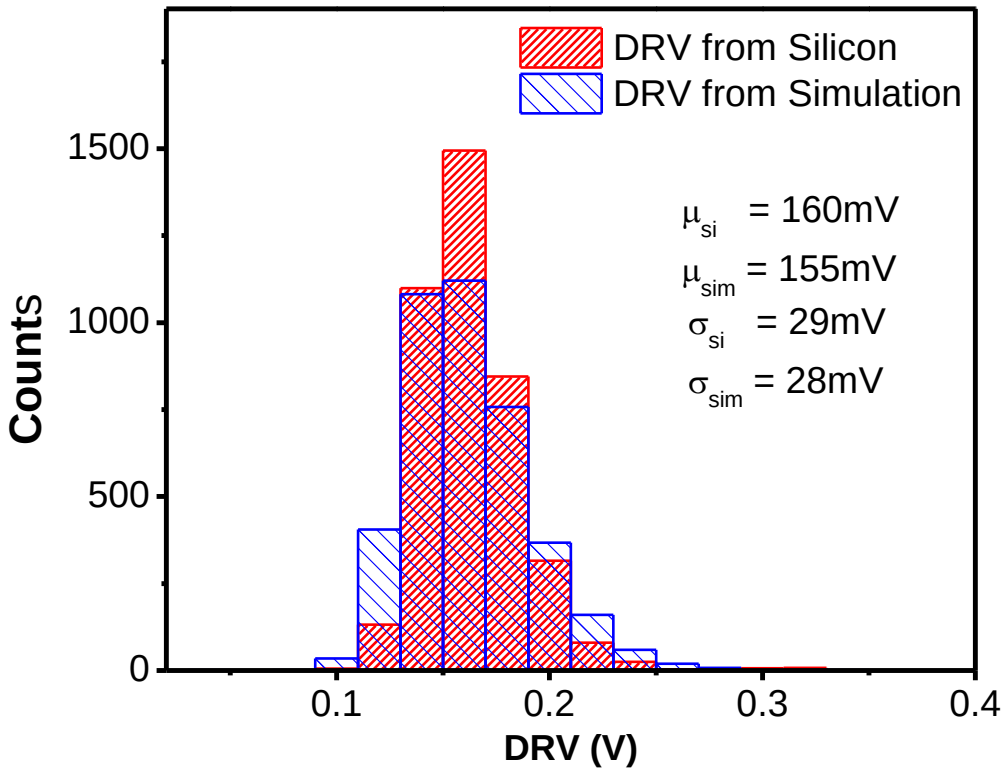

Figure3.4 Validation of simulated DRV using extracted V_{TH} with 90nm silicon data.

CHAPTER 4

AGING: IMPACT ON PERFORMANCE METRICS

4.1 Introduction

Asymmetric scaling of channel length and power supply has resulted in increasing reliability concerns. Use of minimum feature size for SRAM cell design makes it more susceptible to aging as lifetime increases. NBTI (in PMOS devices) and CHC (in NMOS devices) are major reliability concerns in nano-scale regime. In fact, as the gate oxide thickness reaches 4nm, the threshold voltage (V_{th}) degradation of the PMOS transistor caused due to NBTI becomes the limiting factor of the circuit lifetime instead of the channel hot-carrier effect in the NMOS transistor. NBTI occurs when a high voltage is applied at the gate of a PMOS transistor at elevated temperature.

NBTI is a major reliability concern for scaled digital and memory technologies and moreover, as we approach the deep submicron technologies, the manufacturing process reaching its limitations leads to RDF and LER causing variability across devices. In this chapter, the impact of aging and variability is demonstrated leveraging the accurate and efficient V_{TH} extraction methodology presented earlier.

4.2 Device level Aging Statistics

NBTI manifests itself as rise in threshold voltage of PMOS device under the application of negative gate-source bias as mentioned earlier. The fundamental step in circuit aging prediction is to estimate V_{TH} shift in PMOS device under NBTI. For SRAM aging analysis, we use TD (Trapping/De-trapping) based *log(t)* model to predict the shift in mean threshold voltage using extraction methodology. ΔV_{TH} prediction by TD model is written as

$$\Delta V_{th} = \phi\left[A + \log\left(1 + Ct\right)\right] \tag{4.1}$$

where ϕ is proportional to the number of initially available traps per device. The variation in V_{TH} shift is mainly due to ϕ, while parameters A and C are relatively constant. The V_{TH} shift from TD model follows logarithmic relation with time.

It is important to analyze the performance variations of SRAM cell under statistical aging. To investigate the impact of NBTI, we use the extraction methodology to determine the pre-stress and post-stress threshold voltage of PMOS devices in 6T-SRAM cell. Initially the SRAM switch points are measured at high temperature (105^{o}C), which is the temperature set for the aging purpose. SRAM cells on the chip are stressed at V_{DDCELL} = 1.7V at high temperature under static mode. Logic bit stored in the static mode is determined based on the preferred state of the cell by repeated reads. If the preferred state of a cell is ‘logic-0’, the stress condition is set to ‘logic-1’, thus stressing the opposite device. This procedure is illustrated in Figure 4.1. Due to this experiment, one of the PMOS device (P0, P1) experiences threshold voltage shift due to NBTI. After every

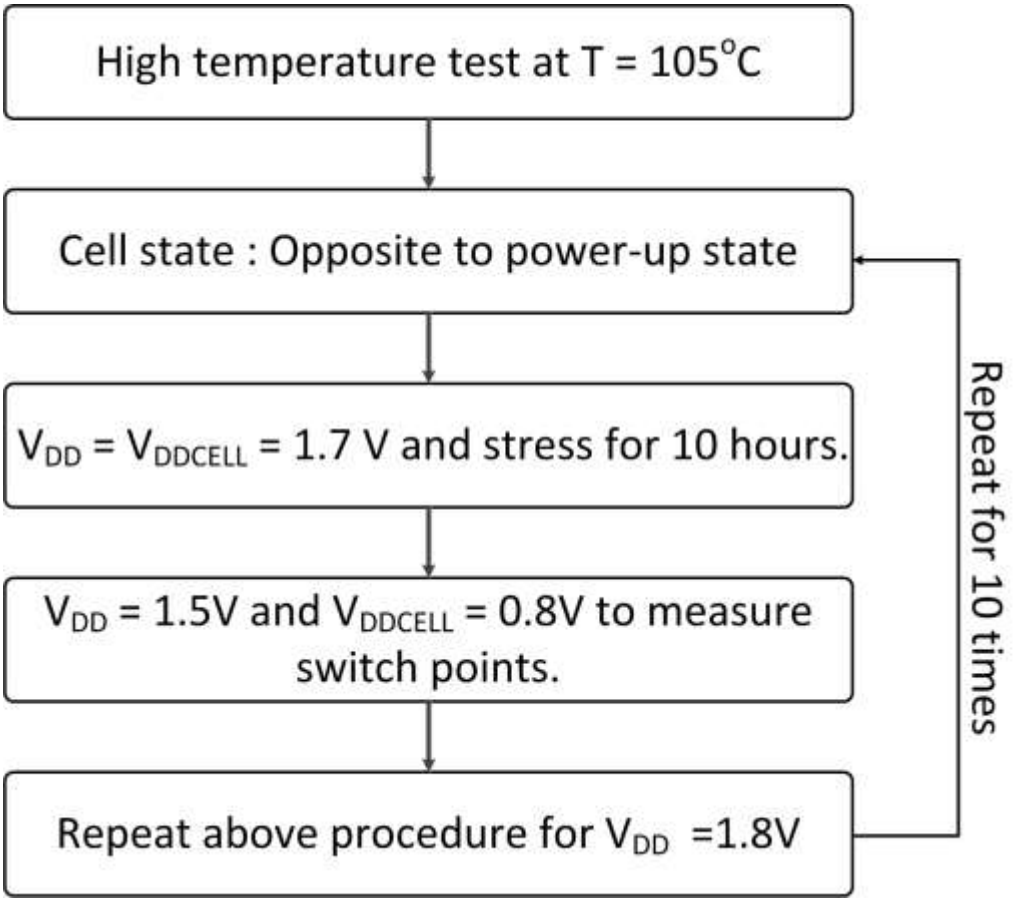

Figure 4.1 Measurement setup followed for aging data measurement from 90nm silicon process

10 hours of stressing, the switch points are collected using the technique described in section 2. This measurement is conducted for 110 hours to collect aging results for SRAM cell array. Following the stress, the V_{TH} extraction procedure is then performed to get threshold voltages of each device at different stress time to analyze the impact of aging.

Figure 4.2 shows the mean shift of threshold voltages in the cross-coupled inverters of SRAM cell. The PMOS V_{TH} increases following the *log(t)* relationship [20-21], and the NMOS V_{TH} is relatively constant under stress, confirming the dominance of NBTI in this technology node. The difference in magnitude of V_{TH} shift in P0 and P1 is

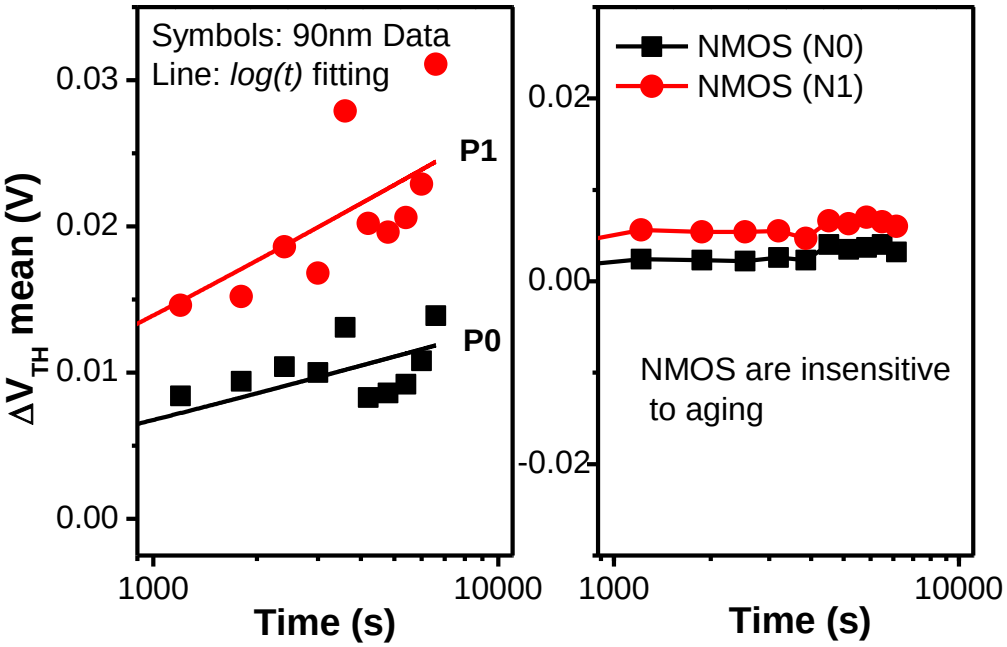

Figure 4.2 Mean of V_{TH} for PMOS and NMOS devices in SRAM cell with increasing stress time. PMOS devices exhibits NBTI degradation following *log(t)* dependence.

due to non-identical stress applied to them. Figure 4.3 presents the extracted V_{TH} of PMOS devices in SRAM cell with increasing stress time. The distribution at the initial time (stress time = 0) suggests the presence of inherent variability due to process variation. As the stress time increases, there is a shift in mean of threshold voltage for both PMOS devices. However, the increase in variance in threshold voltages with the stress time increases slowly and eventually saturates (Figure 4.3(b)). The shift in aging variance is minimal indicating that aging statistics need not be considered while defining guard band at the design stage. With the validation of V_{TH} increase from silicon data, it is Important to study the impact of long-term aging in a 6-T SRAM cell, accounting for both variability and aging.

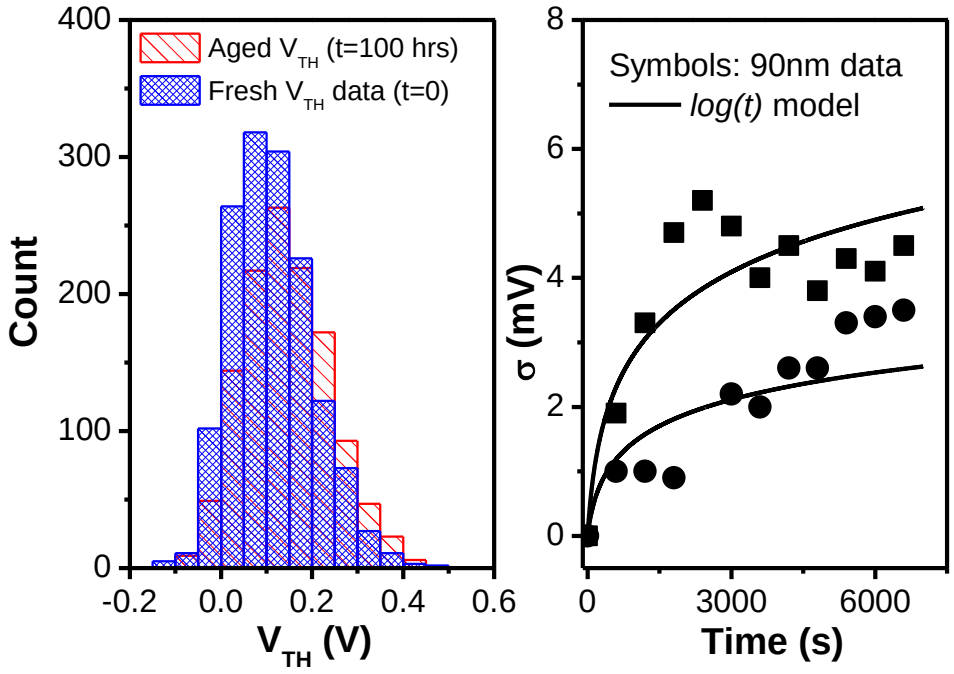

Figure 4.3 V_{TH} distributions for P1 device (Fig. 2) using proposed methodology for increasing stress time.

Table 4.1 Summary of *log(t)* model parameters

Devices	ϕ		A (V)	C (s^{-1})
	μ	σ/μ		
P0	0.0028	26%	1.28e-4	0.0098
P1	0.0059	25%	1.27e-4	0.0099

Apart from V_{TH} mean shift, recent work on NBTI demonstrates the statistical nature of aging [22]. Moreover, Figure 4.4 shows a very weak correlation between the shifts in V_{TH}. It is thus essential to study the impact of statistical aging in SRAM performance indices. In the next section, the

various performance metrics such as static noise margin, read noise margin and data retention voltage under statistical aging is presented.

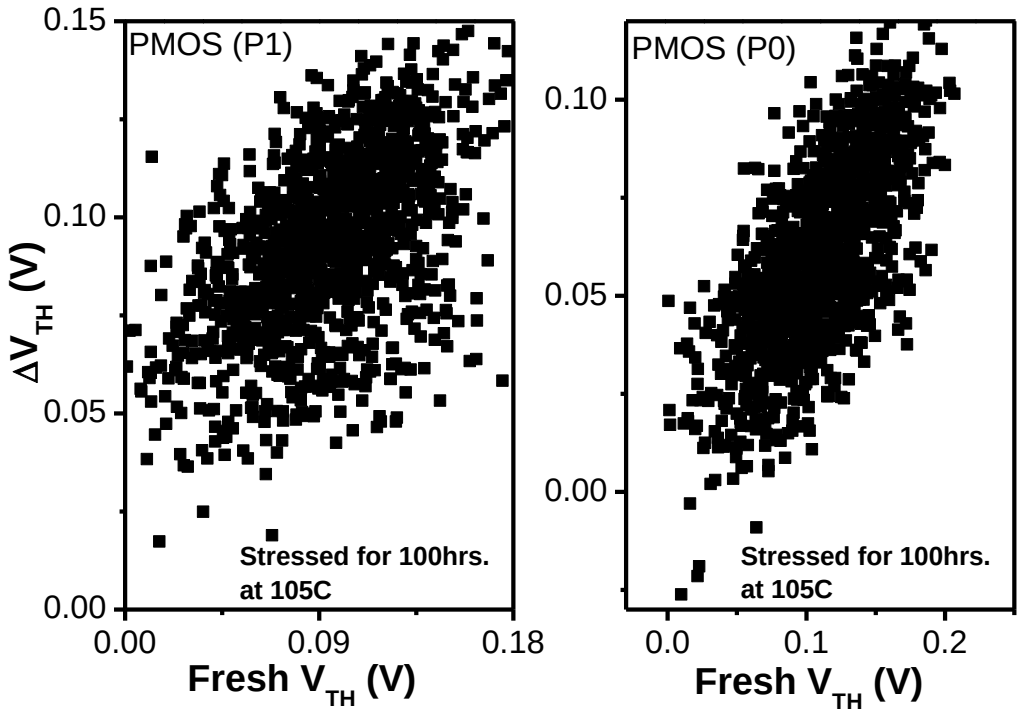

Figure 4.4 Weak correlation between fresh V_{TH} (t=0) and V_{TH} shift..

4.3 Cell Level performance margin analysis

In this section, we study the impact of aging and variability for different design metrics which characterize the performance of a SRAM cell. Static Noise Margin (SNM), Read Noise Margin (RNM) and data retention voltage (DRV) provide effective way of analyzing the stability of SRAM cells. With accurate and efficient V_{TH} extraction methodology at our disposal, we analyze the initial variance affecting the noise margins while; the post-stress extraction further enables statistical aging analysis.

V_{TH} extracted using the extraction methodology is used to predict the initial variance in SRAM cells. Injecting this variance, SRAM cell array is characterized by simulating the static noise margin distribution. This traditional analysis is further extended to determine cell stability under NBTI aging. Under NBTI, threshold voltage of one of the PMOS devices increases since the cell remains in a static state for long time. This systematic shift in V_{TH} is predicted by TD based *log(t)* model which has been discussed in detail in previous section. *log(t)* model is calibrated using silicon measurement, to accurately predict the mean shift in threshold voltage of PMOS devices in 6-T SRAM cell. Based on this model, we estimate the shift in PMOS V_{TH} stress time for two years to analyze the performance degradation. Figure 4.5 shows cumulative distribution function of static noise margin for 4K cells before and after stress. Three different SNM distributions are shown, where fresh cells demonstrate higher SNM. With increment in lifetime, we analyze two different scenarios: mean V_{TH} shift for all the devices and mean V_{TH} shift including the randomness due to aging [21]. For the mean shift for 2 years, there is significant degradation of static noise margin. Further analyzing the impact of aging statistics, variance in aging is compounded with mean V_{TH}

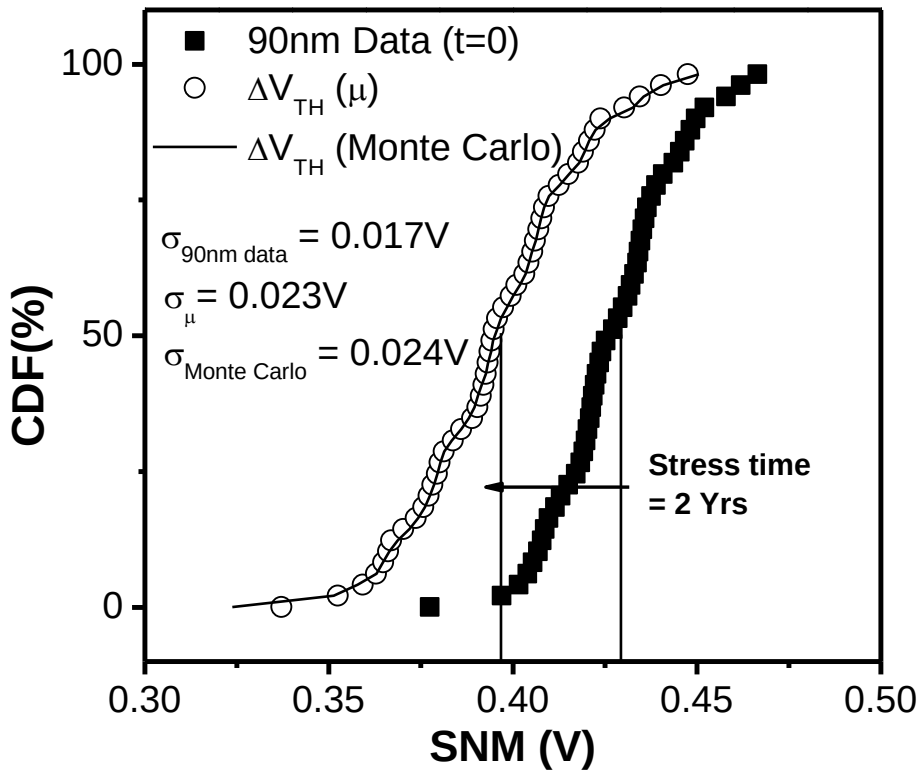

Figure 4.5 Cumulative distribution functions of SNM for fresh and SRAM cell aged for 2 Years

shift. This variance attributed to initial number of traps by TD theory is taken to be 26% of V_{TH} shift [21]. Considering both mean shift and variation in aging, we can see that SNM prediction is similar to the case where only mean shift is considered. Similar analysis is performed for read noise margin and inducing V_{TH} shift in PMOS devices for 2 years, RNM decreases due to degrading cell ratio. Impact of aging variance on RNM distribution is similar to that of RNM (Figure 4.6), indicating that design margins need to address static variability and mean shift due to aging. Additional guar band at the design stage is not required for statistics in aging itself.

Data retention voltage is one of the common performance metrics in SRAM cell, evaluated under aging. Also, DRV being one of the few measureable quantities from existing array structures, memory designers commonly use it to determine guard-band of their design. In our experiment, we study the impact of aging variability on data retention capability of 4K cell array.

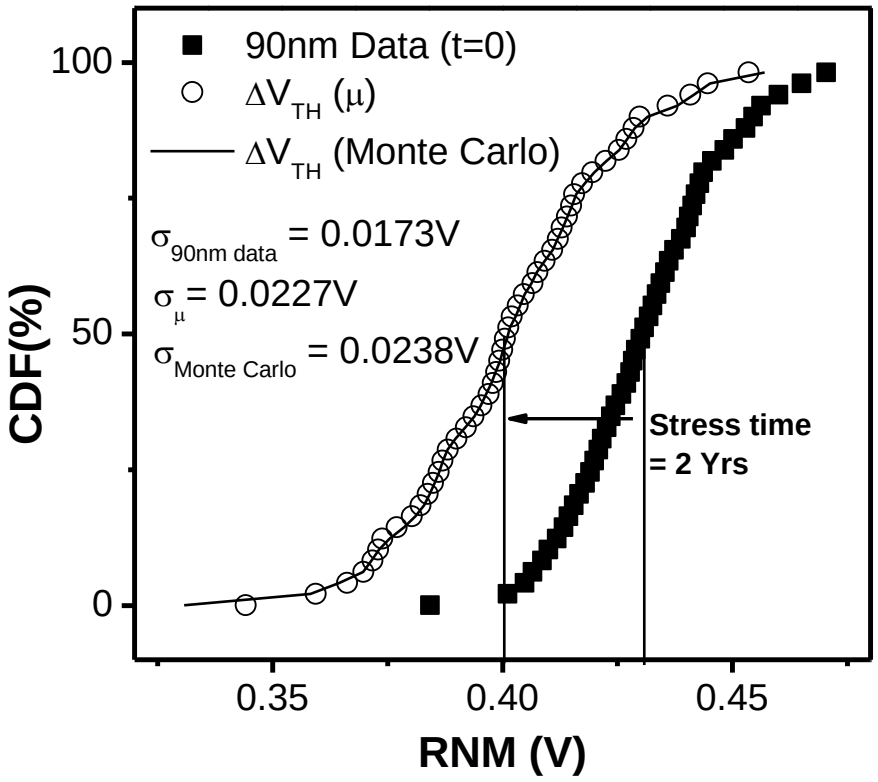

Figure 4.6 Cumulative distribution functions of RNM for fresh and SRAM cell aged for 2 Years

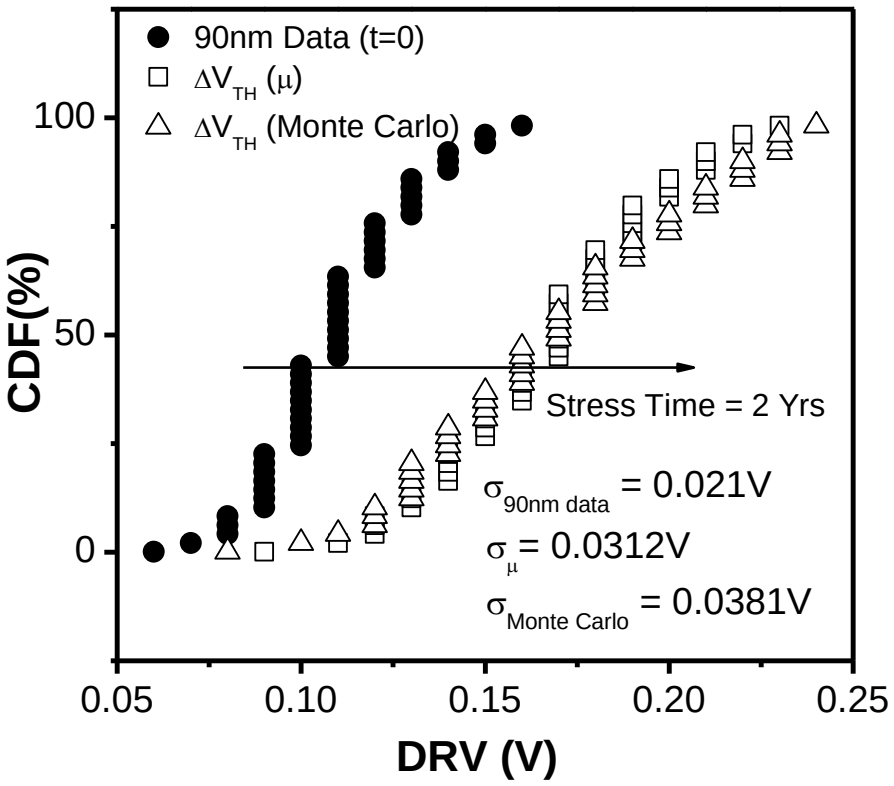

Figure 4.7 Distribution of DRV illustrating the impact of statistical aging.

Using this simulation based approach, we induce the threshold voltage shift and aging variations for 2 years predicted by TD based *log(t)* model as done for SNM and RNM. Simulation based extraction of DRV yields the distribution as presented in Figure 4.6. Statistical Aging of PMOS device in 6T SRAM cell causes a shift in DRV distribution. Aging induced V_{TH} shift demands the increased margins, besides static variability. Thus, impact of static aging and the statistical nature of NBTI have been comprehensively studied in this section. Mean shift in threshold voltage significantly affects the performance metrics such as static noise margin, read margin and the data retention voltage. Although, variance due to aging does not have a severe impact, mean shift in performance metrics and static variations define margins during the design stage.

4.4 Aging: Compensation of Mismatch in V_{TH}

In general, the static SRAM memories are not accessed so frequently and they remain in static state for considerable amount of time. Depending on whether a logic '1' or logic '0' is stored in the cell, one of the PMOS device is stressed during this idle time. This is the aging impact as pointed out earlier affecting all the performance metrics. We also pointed out that due to smallest geometries adopted for SRAM array for the reason to design dense memories; the impact of variability due to technology scaling has become very critical. In this work, we tried to leverage this aging to our advantage to compensate for the mismatch in V_{TH} between devices by allowing the device to age and thereby increase the threshold voltage [18].

Figure 4.8 shows the correlation map of the 8 switch points between fresh and aged measurements. The randomness clearly states an important pointer that there inherent variation present in aging itself which adds on the top already present variation and hence this approach is not feasible for our purpose.

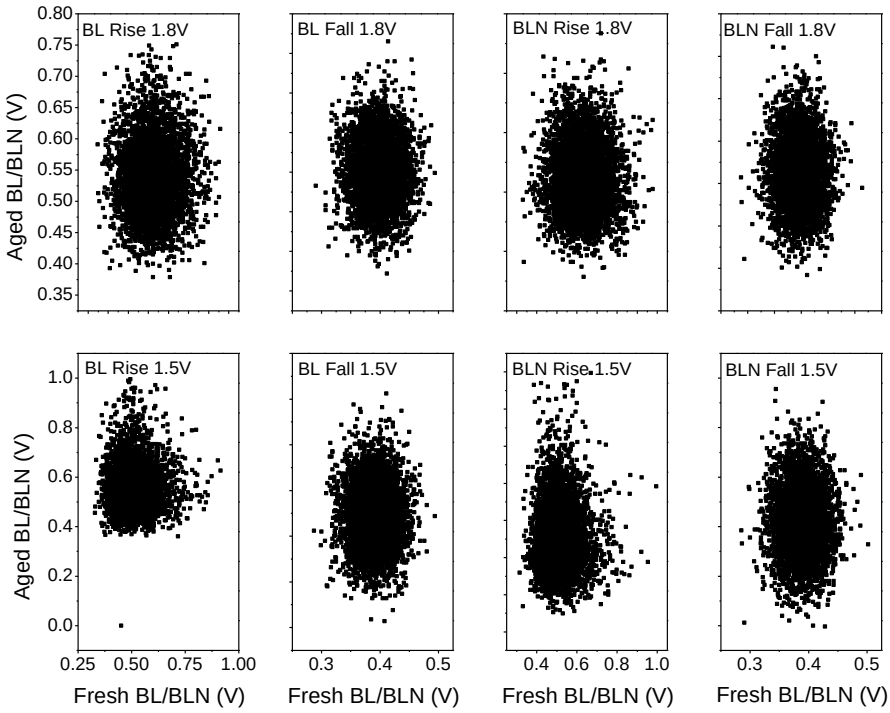

Figure 4.8 Correlation map between fresh and aged switch points.

CHAPTER 5

SUMMARY AND FUTURE SCOPE OF WORK

5.1 Thesis Conclusions

In this work towards my Master's thesis, I have covered the following topics relating to the reliability and variability issues related to SRAM memory arrays. SRAM cell design is beginning to become a challenging and critical game in today's CMOS technology due to the aggressive scaling and low geometries. I have discussed on the various causes of reliability or variability concerns and also presented a method to efficient extract the V_{TH} due to this variability for each device in a SRAM cell, keeping with a non-invasive silicon setup not affecting the normal memory operation. The specific contributions are as listed below.

- Chapter 1 gives an introduction to the criticality of reliability and variability concerns in today's nano-scale technology regime. It also talks about the role of technology scaling on reliability and yield and also gives insight on the role of memory and how it is impacted by variability concerns.
- In Chapter 2 starts with an outline of the test principle and draws into the methodology presented to extracted the V_{TH} of each individual device in the SRAM cell. It presents the details models for transistors along with the concept behind the extraction procedure which is based on identifying the operating region of each device in the cell.
- A through validation is covered in Chapter 3 starting from a simulation framework and working its way into silicon validation.
- Chapter 4 introduces the concept of circuit reliability due to aging. It talks about the various aging mechanisms but focusing on NBTI which is a dominant factor in today's digital

circuits. It also presents the *log(t)* model used to predict the aging in the PMOS device along with how the various performance metrics are altered due to this.

5.2 Future Work

The work from this thesis indicates the importance of aging statistics and variability in nano-scale technologies. The results clearly indicate the areas where guard-banding of the design can be done. However, there is still a lot of study to be done on the impact of aging and variability in advanced process nodes. There are various variants of SRAM cell 8-T and 10-T to which this work can be extended to. Also in addition to that, similar analysis can be extended to other types of memory like DRAM, MRAM, PCRAM to study the impact of variations and aging on them.

REFERENCES

[1] Lidia Łukasiak and Andrzej Jakubowski, “History of Semiconductors”, Invited paper, Journal of Telecommunications and Information Technology. 2010.

[2] Gordon E. Moore, “Cramming More Components onto Integrated Circuits,” Electronics, pp. 114‑117, April 19, 1965.

[3] “Digital Integrated Circuits”, Jan M. Rabaey, Anantha Chandrakasan, Borivoje Nikolic. Second Edition.

[4] “CMOS VLSI Design”, Neil H.E.Weste, David Harris, Third Edition.

[5] Chi-Chao Wang et al., “Modeling of Layout-Dependent Stress Effect in CMOS Design,” ICCAD, 2009.

[6] Anthony.S. Oates, “Reliability Issues for High-K Gate Dielectrics”,IEDM 2003.

[7] http://en.wikipedia.org/wiki/Random_dopant_fluctuation

[8] J.A. Croon et al., “Line Edge Roughness: Characterization, Modeling and Impact on Device Behavior,” IEDM, 2002.

[9] Jeffrey Hicks et al., “45nm Transistor Reliability,”, Intel Technology Journal, Volume 12, Issue 2, 2008.

[10] S. Pae, J. Maiz, C. Prasad, and B. Woolery, “Effect of BTI degradation on transistor variability in advanced semiconductor technologies,” IEEE Trans. Dev. Mat. Rel., vol. 8, no. 3, pp. 519‑525, Sep. 2008.

[11] B. Vaidyanathan and A. S. Oates, “Technology scaling effect on the relative impact of transistor degradation and process variation on the re-liability of digital circuits,” IEEE Trans. Dev. Mat. Rel., to be published.

[12] S. Chellappa, et al. “In-situ Characterization and Extraction of SRAM variability,” DAC, pp. 711-716, 2010.

[13] R. Vattikonda, W. Wang, and Y. Cao, “Modeling and minimiza-tion of pmos nbti effect for robust nanometer design”, DAC, pp. 1047‑1052, Jul. 2006.

[14] B. Mauck, V. Ravichandran, and U. Mughal, "A design for test technique for parametric analysis of SRAM: On-die low yield analysis," Int. Test Conf., pp. 105-113, Feb. 2004.

[15] N. Gierczynski, B. Borot, N. Planes and H. Brut, "A new combined methodology for write-margin extraction of advanced SRAM," IEEE Int. Conf. on Microelectronic Test Structures, pp. 97-100, Mar. 2007.

[16] K. Takeda, et al, "Redefinition of write margin for next generation SRAM and write-margin monitoring circuit," ISSCC, pp. 630‐631, Feb. 2006.

[17] E. Grossar, M. Stucchi, K. Maex and W. Dehaene, "Read stability and write-ability analysis of SRAM cells for nanometer technologies," IEEE JSSC, vol. 41, no.11, pp.2577-2587, Nov. 2006

[18] Venkatesa Ravi et al., "Statistical Aging Analysis and Prediction of SRAM Cell Margins", submitted to DAC 2013.

[19] T. Mizutani, A. Kumar, T. Hiramoto, "Measuring threshold voltage variability of 10G transistors," IEDM , pp.25.2.1-25.2.4, 2011.

[20] R. da Silva and G. Wirth, "Logarithmic behavior of the degradation dynamics of metal oxide semiconductor devices", Journal of Statistical Mechanics, vol. P04035, pp. 1-12, 2010.

[21] J. Velamala, K. Sutaria, T. Sato and Y. Cao, "Aging statistics based on trapping/detrapping: Silicon evidence, modeling and long-term prediction," IRPS, 2F.2.1-2F.2.5, 2012.

[22] S. Pae, J. Maiz and C. Prasad, "Effect of NBTI degradation on transistor variability in advanced technologies," Integrated Reliability Workshop, vol., no., pp.18-21, 15-18 Oct. 2007.

[23] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," J. Applied Physics, vol. 94, no. 1, pp. 1-18, 2003.

[24] The International Technology Roadmap for Semiconductors (ITRS), 2009.

APPENDIX A

KCL EQUATION SOLVED

Note: The KCL and solving algorithm is explained for the RISE case signal at input.

$$I_{N0} = I_{NA0} \tag{1}$$

$$I_{N1} = I_{P1} \tag{2}$$

From (2)

$$A_p e^{[(V_{DD}-V_D)-\gamma_p V_{thp1}+\theta_p (V_{DD}-V_Q)]/d_p} = A_n e^{[(V_D)-\gamma_n V_{thn1}+\theta_n (V_Q)]/d_n}$$

Taking natural log on either sides.

$$\ln(A_p) + \frac{[(V_{DD}-V_D)-\gamma_p V_{thp1}+\theta_p (V_{DD}-V_Q)]}{d_p} = \ln(A_n) + \frac{[(V_D)-\gamma_n V_{thn1}+\theta_n (V_Q)]}{d_n}$$

Re-arranging to group terms to represent V_D,

$$V_D = \frac{[\ln(A_p) - \ln(A_n) - \gamma_p V_{thp1} + \gamma_n V_{thn1} - \theta_n (V_Q) + \theta_p (V_{DD} - V_Q) + (V_{DD}/d_p)]}{(d_n + d_p)}$$

From (1),

$$K_{na}[(V_{WL} - \beta V_{BL} - \lambda_n V_{thna0})(V_D - V_{BL})] = K_n[V_D - \eta_n V_{thn0} + \alpha_n V_Q]$$

Substituting the expression for V_D from above into this equation we reduce the system with just V_{TH} as unknowns.

In the fall case, the KCL has P-channel device in place of N0 device.

APPENDIX B

MATLAB CODE: FITTING IV AND SOLVER

```
clear all;
close all;
clc
 data=xlsread('1130\Final_aged_range_Vbl_datapoints.xlsx');
A = [ 0.5 ; 0.5 ; 0.5 ; 0.5 ;0.5 ; 0.5 ] ;
[DN_RISE_p1,GN_RISE_p1,TN_RISE_p1,AN_RISE_p1,DP_RISE_p1,GP_RISE_p1,TP_RISE
_p1,AP_RISE_p1,DN_FALL_p1,GN_FALL_p1,TN_FALL_p1,AN_FALL_p1,DP_FALL_p1,G
P_FALL_p1,TP_FALL_p1,AP_FALL_p1,KNAL_RISE15_p1,BL_RISE15_p1,GL_RISE15_p1,
KNAL_RISE18_p1,BL_RISE18_p1,GL_RISE18_p1,KNAR_RISE15_p1,BR_RISE15_p1,GR_
RISE15_p1,KNAR_RISE18_p1,BR_RISE18_p1,GR_RISE18_p1,KNAL_FALL15_p1,BL_FAL
L15_p1,GL_FALL15_p1,KNAL_FALL18_p1,BL_FALL18_p1,GL_FALL18_p1,KNAR_FALL
15_p1,BR_FALL15_p1,GR_FALL15_p1,KNAR_FALL18_p1,BR_FALL18_p1,GR_FALL18_
p1,KNL15_p1,ANL15_p1,GNL15_p1,KNL18_p1,ANL18_p1,GNL18_p1,KPL15_p1,APL15_p
1,GPL15_p1,KPL18_p1,APL18_p1,GPL18_p1,KNR15_p1,ANR15_p1,GNR15_p1,KNR18_p1,
ANR18_p1,GNR18_p1,KPR15_p1,APR15_p1,GPR15_p1,KPR18_p1,APR18_p1,GPR18_p1] =
Fit_IV_test_p1 ();
L=size(data);
for i = 1:1:L(2)
Vblr15 = data(6,i);    Vblr18 = data(2,i);    Vblf15 = data(7,i);    Vblf18 =  data(3,i);    Vbrr15 =
data(8,i);    Vbrr18 = data(4,i);    Vbrf15 = data(9,i);   Vbrf18 = data(5,i);
```

```
%g =
@(Y)Solve_test(Y,Vblr15,Vblr18,Vblf15,Vblf18,Vbrr15,Vbrr18,Vbrf15,Vbrf18,DN_RISE,GN_
RISE,TN_RISE,AN_RISE,DN_FALL,GN_FALL,TN_FALL,AN_FALL,DP_RISE,GP_RISE,T
P_RISE,AP_RISE,DP_FALL,GP_FALL,TP_FALL,AP_FALL,KNAL_RISE15,BL_RISE15,GL
_RISE15,KNAL_RISE18,BL_RISE18,GL_RISE18,KNAR_RISE15,BR_RISE15,GR_RISE15,
KNAR_RISE18,BR_RISE18,GR_RISE18,KNAL_FALL15,BL_FALL15,GL_FALL15,KNAL_
FALL18,BL_FALL18,GL_FALL18,KNAR_FALL15,BR_FALL15,GR_FALL15,KNAR_FAL
L18,BR_FALL18,GR_FALL18,KNL15,ANL15,GNL15,KNL18,ANL18,GNL18,KPL15,APL15
,GPL15,KPL18,APL18,GPL18,KNR15,ANR15,GNR15,KNR18,ANR18,GNR18,KPR15,APR1
5,GPR15,KPR18,APR18,GPR18);
 [X(:,i),Ssq,CNT,Res,xy] = LMFnlsq(g,A,'Display',1,'Xtol',1e-6,'FunTol',1e-
6,'MaxIter',1000,'Trace', 0,'Printf' , 'printit');
```

APPENDIX C

MATLAB CODE: EQUATIONS

```
function f =
Solve_test(Y,Vblr15,Vblr18,Vblf15,Vblf18,Vbrr15,Vbrr18,Vbrf15,Vbrf18,DN_RISE,GN_RISE
,TN_RISE,AN_RISE,DN_FALL,GN_FALL,TN_FALL,AN_FALL,DP_RISE,GP_RISE,TP_RI
SE,AP_RISE,DP_FALL,GP_FALL,TP_FALL,AP_FALL,KNAL_RISE15,BL_RISE15,GL_RIS
E15,KNAL_RISE18,BL_RISE18,GL_RISE18,KNAR_RISE15,BR_RISE15,GR_RISE15,KNA
R_RISE18,BR_RISE18,GR_RISE18,KNAL_FALL15,BL_FALL15,GL_FALL15,KNAL_FAL
L18,BL_FALL18,GL_FALL18,KNAR_FALL15,BR_FALL15,GR_FALL15,KNAR_FALL18,
BR_FALL18,GR_FALL18,KNL15,ANL15,GNL15,KNL18,ANL18,GNL18,KPL15,APL15,GP
L15,KPL18,APL18,GPL18,KNR15,ANR15,GNR15,KNR18,ANR18,GNR18,KPR15,APR15,G
PR15,KPR18,APR18,GPR18)
    dn_rise = DN_RISE;
    gn_rise = GN_RISE;
    tn_rise = TN_RISE;
    an_rise = AN_RISE;
    dn_fall = DN_FALL;
    gn_fall = GN_FALL;
    tn_fall = TN_FALL;
    an_fall = AN_FALL;
    %% P1 transistor %%

    dp_rise = DP_RISE ;
    gp_rise = GP_RISE;
```

```
ap_rise = AP_RISE;

tp_rise = TP_RISE;

dp_fall = DP_FALL;

gp_fall = GP_FALL;

tp_fall = TP_FALL;

ap_fall = AP_FALL;

%% N1 %%

dn_fall = DN_FALL;

gn_fall = GN_FALL;

tn_fall = TN_FALL;

an_fall = AN_FALL;

%% P1 %%

dp_fall = DP_FALL;

gp_fall = GP_FALL;

tp_fall = TP_FALL;

ap_fall = AP_FALL;

%%%%%%%% Access transistor %%%%%%%

%% LHS 1.5 Rise

knal_rise15 = KNAL_RISE15;

bl_rise15 = BL_RISE15;

gl_rise15 = GL_RISE15;

%% LHS 1.8 Rise
```

```
knal_rise18 = KNAL_RISE18;

bl_rise18 = BL_RISE18;

gl_rise18 = GL_RISE18;

%% RHS 1.5 Rise

knar_rise15 =KNAR_RISE15;

br_rise15 = BR_RISE15;

gr_rise15 = GR_RISE15;

%% RHS 1.8 Rise

knar_rise18 =KNAR_RISE18;

br_rise18 = BR_RISE18;

gr_rise18 = GR_RISE18;

%% LHS 1.5 Fall

knal_fall15 = KNAL_FALL15;

bl_fall15 = BL_FALL15;

gl_fall15 = GL_FALL15;

%% LHS 1.8 Fall

knal_fall18 = KNAL_FALL18;

bl_fall18 = BL_FALL18;

gl_fall18 = GL_FALL18;

%% RHS 1.5 Fall

knar_fall15 =KNAR_FALL15;

br_fall15 = BR_FALL15;
```

```
gr_fall15 = GR_FALL15;
%% RHS 1.8 Fall
knar_fall18 =KNAR_FALL18;
br_fall18 = BR_FALL18;
gr_fall18 = GR_FALL18;
%% N0 %%
knl15=KNL15;
anl15=ANL15;
gnl15=GNL15;
knl18=KNL18;
anl18=ANL18;
gnl18=GNL18;
knr15=KNR15;
anr15=ANR15;
gnr15=GNR15;
knr18=KNR18;
anr18=ANR18;
gnr18=GNR18;
%% P0 %%
kpl15=KPL15;
apl15=APL15;
gpl15=GPL15;
```

```
kpl18=KPL18;

apl18=APL18;

gpl18=GPL18;

kpr15=KPR15;

apr15=APR15;

gpr15=GPR15;

kpr18=KPR18;

apr18=APR18;

gpr18=GPR18;

%% Vbl mapping%%

vblr15 = Vblr15;

vblr18 = Vblr18;

vblf15 = Vblf15;

vblf18 = Vblf18;

vbrr15 = Vbrr15;

vbrr18 = Vbrr18;

vbrf15 = Vbrf15;

vbrf18 = Vbrf18;

%% Thresholds Defined %%

vqr = 0.7;

vqf = 0.2;

%C1 - LHS Rise 1.5
```

```
C1 = (((ap_rise + 0.8*dp_rise - (gp_rise*Y(2,1)) + (tp_rise*(0.8-vqr)) - an_rise +
(gn_rise*Y(4,1)) - tn_rise*vqr) / (dn_rise +dp_rise)));
%C2 - LHS Rise 1.8
C2 = (((ap_rise + 0.8*dp_rise - (gp_rise*Y(2,1)) + (tp_rise*(0.8-vqr)) - an_rise +
(gn_rise*Y(4,1)) - tn_rise*vqr) / (dn_rise +dp_rise)));
%C3 - LHS Fall 1.5
C3 = (((ap_fall + 0.8*dp_fall - (gp_fall*Y(2,1)) + (tp_fall*(0.8-vqf)) - an_fall +
(gn_fall*Y(4,1)) - tn_fall*vqf) / (dn_fall +dp_fall)));
%C4 - LHS Fall 1.8
C4 = (((ap_fall + 0.8*dp_fall - (gp_fall*Y(2,1)) + (tp_fall*(0.8-vqf)) - an_fall +
(gn_fall*Y(4,1)) - tn_fall*vqf) / (dn_fall +dp_fall)));
%C5 RHS Rise 1.5
C5 = (((ap_rise + 0.8*dp_rise - (gp_rise*Y(1,1)) + (tp_rise*(0.8-vqr)) - an_rise +
(gn_rise*Y(3,1)) - tn_rise*vqr) / (dn_rise +dp_rise)));
%C6 RHS Rise 1.8
C6 = (((ap_rise + 0.8*dp_rise - (gp_rise*Y(1,1)) + (tp_rise*(0.8-vqr)) - an_rise +
(gn_rise*Y(3,1)) - tn_rise*vqr) / (dn_rise +dp_rise)));
%C7 RHS Fall 1.5
C7 = (((ap_fall + 0.8*dp_fall - (gp_fall*Y(1,1)) + (tp_fall*(0.8-vqf)) - an_fall +
(gn_fall*Y(3,1)) - tn_fall*vqf) / (dn_fall +dp_fall)));
%C8 RHS Fall 1.8
```

```
C8 = (((ap_fall + 0.8*dp_fall - (gp_fall*Y(1,1)) + (tp_fall*(0.8-vqf)) - an_fall +
(gn_fall*Y(3,1)) - tn_fall*vqf) / (dn_fall +dp_fall)));
f = [%LHS Rise 1.5
((knal_rise15*(1.5 - (bl_rise15*C1) - (gl_rise15*Y(5,1)))*(vblr15 - C1)) - (knl15*(vqr -
(gnl15*Y(3,1)) + anl15*C1)));
% LHS Rise 1.8
((knal_rise18*(1.8 - (bl_rise18*C2) - (gl_rise18*Y(5,1)))*(vblr18 - C2)) - (knl18*(vqr -
(gnl18*Y(3,1)) + anl18*C2)));
%LHS Fall 1.5
((knal_fall15*(1.5 - (bl_fall15*vblf15) - (gl_fall15*Y(5,1)))*(C3 - vblf15)) -
(kpl15*(0.8 - vqf - (gpl15*Y(1,1)) + apl15*(0.8-C3)))) ;
%LHS Fall 1.8
((knal_fall18*(1.8 - (bl_fall18*vblf18) - (gl_fall18*Y(5,1)))*(C4- vblf18)) -
(kpl18*(0.8 - vqf - (gpl18*Y(1,1)) + apl18*(0.8-C4)))) ;
%RHS Rise 1.5
((knar_rise15*(1.5 - (br_rise15*C5) - (gr_rise15*Y(6,1)))*(vbrr15 - C5)) - (knr15*(vqr
- (gnr15*Y(4,1)) + anr15*C5))) ;
%RHS Rise 1.8;
((knar_rise18*(1.8 - (br_rise18*C6) - (gr_rise18*Y(6,1)))*(vbrr18 - C5)) - (knr18*(vqr
- (gnr18*Y(4,1)) + anr18*C6))) ;
%RHS Fall 1.5
```

((knar_fall15*(1.5 - (br_fall15*vbrf15) - (gr_fall15*Y(6,1)))*(C7 - vbrf15)) - (kpr15*(0.8 - vqf - (gpr15*Y(2,1)) + apr15*(0.8-C7)))) ;

% RHS Fall 1.8

((knar_fall18*(1.8 - (br_fall18*vbrf18) - (gr_fall18*Y(6,1)))*(C8 - vbrf18)) - (kpr18*(0.8 - vqf - (gr18*Y(2,1)) + apr18*(0.8-C8)))) ;];