Reliability Modeling and Mitigation for Embedded Memories

Agbo, Innocent

DOI
10.4233/uuid:ce7b3290-9e0f-406b-93ee-7bfb7c9a8430

Publication date
2018

Document Version
Final published version

Citation (APA)
Agbo, I. (2018). *Reliability Modeling and Mitigation for Embedded Memories*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:ce7b3290-9e0f-406b-93ee-7bfb7c9a8430

RELIABILITY MODELING AND MITIGATION FOR EMBEDDED MEMORIES

RELIABILITY MODELING AND MITIGATION FOR EMBEDDED MEMORIES

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op maandag 18 June 2018 om 15:00 uur

door

Innocent Okwudili AGBO

Master of Science in Computer Engineering,
Technische Universiteit Delft, Delft, Nederland,
geboren te Amechi-Awkunanaw, Nigeria.

Dit proefschrift is goedgekeurd door de

promotor:
Prof. dr. ir. S. Hamdioui

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. S. Hamdioui	Technische Universiteit Delft, promotor
Onafhankelijke leden:	
Prof. dr. ir. K. L. M. Bertels	Technische Universiteit Delft
Prof. dr. G. Q. Zhang	Technische Universiteit Delft
Prof. dr. F. Catthoor	Katholieke Universiteit Leuven, Belgium
Prof. dr. P. Girard	Laboratoire d'Informatique, de Robotique et de Microelectronique de Montpellier, France
Dr. ir. A. C. Kruseman	NXP Semiconductors, Eindhoven, Nederland
Prof. dr. ir. G. J. T. Leus	Technische Universiteit Delft, reservelid
Overig lid:	
Dr. ir. M. Taouil	Technische Universiteit Delft

Keywords: Memory reliability, Aging, Bias Temperature Instability (BTI), Sense Amplifier.

Printed by: IPSKAMP printing, Enschede, the Netherlands.

Front & Back: Beautiful cover art that captures the entire content of this thesis in a single illustration is taken from http://keas-group.com/expertise/d-fabrication-de-cartes-electroniques-en/

 ioagbo@gmail.com

ISBN/EAN: 978-94-6366-053-2

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Dedicated to the memory of my father
Mr. Michael Onyianwugwu Agbo.

SUMMARY

Complementary Metallic Oxide Semiconductor (CMOS) technology scaling enhances the performance, transistor density, functionality, and reduces cost and power consumption. However, scaling causes significant reliability challenges both from a manufacturing and operational point of view. Obtaining reliable memories require accurate understanding of the impact of aging (*such as Bias temperature instability (BTI)*) on individual memory components and how they interact with each other. In this dissertation, two types of challenges are addressed, which are related to BTI aging and partially to mitigation schemes: one related to the aging of sense amplifier and another one to the aging of read path and write path.

Analysis of aging impact on different memory sense amplifiers - The analysis of BTI impact on various memory sense amplifier (SA) designs was performed, while taking into account two BTI models (i.e., Atomistic and RD model), different technology nodes (i.e., 90, 65, 45, 32, 22, and 16 nm), and different workloads. First, the analysis and comparison of RD and Atomistic models impact on the SA were performed. The results show that the atomistic trap-based BTI model is more accurate than the RD model. Second, the investigation of BTI impact on the drain-input latch type SA for various technology nodes and supply voltages was performed. The result shows that as technology scales down, the impact of BTI on sensing delay increases, while the sensing voltage decreases, causing less robust and reliable memory sense amplifier. The result also shows that increase in supply voltage compensates the BTI degradation. Third, an accurate technique was proposed and characterized for the integral impact of BTI and voltage temperature variation on the memory standard latch type SA for various technology nodes and workloads. The results show that the degradation is strongly dependent on workload and temperature. Fourth, in addition to the latter, the impact of process variation at time-zero was incorporated and analyzed. The results show that the SA sensing delay degradation is more significant at lower nodes and could lead to read failures at lower power supply. This reveals that there must be a tradeoff between performance and reliability. Fifth, an accurate methodology was proposed to quantify the impact of variability on the memory SA offset-voltage for both time-zero and time-dependent variability. The results show that the impact on the offset voltage specification is significant for aging time-dependent variability. Sixth, on top of the latter, the sensitivity of the SA and its failure rate were analyzed for five process corners (i.e., Nominal, Fast-Fast, Fast-Slow, Slow-Fast, and Slow-Slow). The results show that balanced workloads result in a significant low offset voltage specification. Finally, the impact of aging was analyzed and compared, while considering different supply voltages, temperatures, and SA designs. The results show that the High Performance SA degrades faster than other SA types, irrespective of the workload, supply voltage, and temperature.

Investigation of read path aging - Adequate techniques was proposed to estimate and mitigate the impact of aging on the read path of a high performance SRAM memory. The mitigation techniques are based on the re-sizing of the pull-down transistors of the cell's and the SA's designs. The results show that the SA mitigation is more effective for the SRAM read path (i.e., SA) than cell mitigation.

Investigation of write path aging - The analysis of BTI impact on the SRAM write driver was performed for various supply voltages, temperatures, and technology nodes. The result shows that the impact of BTI increases the write delay and widen its distribution, when the technology scales down.

SAMENVATTING

Complementaire-metaaloxide-halfgeleider-technologieverkleining verbetert de prestaties, transistordichtheid, functionaliteit en reduceert de kosten en het energieverbruik. Echter, verkleining veroorzaakt significante betrouwbaarheidsproblemen, zowel vanuit productie- als operationeel oogpunt. Het verkrijgen van betrouwbare geheugens vereist een goed begrip van het effect van veroudering (zoals bias temperatuur instabiliteit (BTI)) op individuele geheugencomponenten en hoe ze elkaar wederzijds beïnvloeden. In dit proefschrift worden twee typen uitdagingen geadresseerd, welke gerelateerd zijn aan BTI veroudering en gedeeltelijk aan verlichtingsmethodes: één gerelateerd aan de veroudering van de geheugenleesversterker, de andere aan de veroudering van het leespad en schrijfpad.

Analyse van het effect van veroudering op verschillende leeversterkerontwerpen - De analyse van het effect van BTI op verschillende geheugenleesversterkerontwerpen is uitgevoerd terwijl twee BTI modellen in acht zijn genomen (i.e., atomistische en RD model), verschillende technologienodes (i.e., 90, 65, 45, 32, 22 en 16 nm), en verschillende belastingen. Ten eerste zijn de analyse en vergelijking van de invloed van de RD en atomistische modellen op de geheugenleesversterker uitgevoerd. De resultaten tonen aan dat het atomistiche val-gebassseerde BTI model accurater is dan het RD model. Ten tweede is de analyse van het BTI effect op de drain-input latch type geheugenleesversterker uitgevoerd voor verschillende technologienodes en voedingsspaningen. De resultaten tonen dat het effect van BTI op de leesvertraging toeneemt en de leesspanning afneemt met technologieverkleining, wat leidt tot een minder robuuste en betrouwbare geheugenleesversterker. De resultaten tonen ook dat een toename in de voedingsspaning compenseert voor BTI degradatie. Ten derde, is een accurate methode voorgedragen, welke het integrale effect van BTI en spannings- en temperatuurvariaties op de standaard latch type geheugenleesversterker karakteriseert voor verschillende technologienodes en belastingen. De resultaten tonen dat de degradatie sterk afhangt van de belasting en temperatuur. Ten vierde, als toevoeging op het laatstgenoemde, is het effect van procesvariatie toegevoegd en geanalyseerd. De resultaten tonen dat de degradatie van de leesvertraging van de geheugenleesversterker significanter is op kleinere nodes en dat dit zou kunnen leiden tot leesfouten op lage voedingsspaningen. Dit onthult dat er een afweging is tussen prestaties en betrouwbaarheid. Ten vijfde, is een accurate methodologie voorgesteld om het effect op de offset-spanning van geheugenleesversterkers te kwantificeren voor zowel tijd-nul en tijdsafhankelijke variabiliteit. De resultaten tonen dat het effect van tijdsafhankelijke variabiliteit aanzienlijk is op de offset-spanningsspecificatie. Ten zesde, als toevoeging op het laatstgenoemde, is de gevoeligheid van de geheugenleesversterker en zijn uitvalspercentage voor vijf procédé-uitersten (i.e., Nominaal, Snel-Snel, Snel-Traag, Traag-Snel en Traag-Traag) onderzocht. De resultaten tonen dat gebalanceerde belastingen resulteren in een significant lagere offset-

spanningsspecificatie. Als laatste, is het effect van veroudering geanalyseerd en vergeleken, terwijl verschillende voedingsspanningen, temperaturen en leesversterkerontwerpen in acht zijn genomen. De resultaten tonen dat de hoge-prestatieleesversterker sneller degradeert dan andere typen leesversterkers, onafhankelijk van belasting, voedingsspaning en temperatuur.

Analyse van leespadveroudering – Bekwame technieken zijn voorgedragen om het effect van veroudering op het leespad van een hoge-prestatie-SRAM-geheugen te schatten en verlichten. De verlichtingstechnieken zijn gebasseerd op het herschalen van de pull-down transistoren van de geheugenelement- en leesversterkerontwerpen. De resulaten tonen dat geheugenleesversterkerverlichting effectiever is voor het SRAM-leespad dan geheugenelementverlichting.

Analyse van schrijfpadveroudering – De analyse van het effect van BTI op de SRAM-schrijfaandrijver voor verschillende voedingsspanningen, temperaturen en technologienodes is uitgevoerd. De resultaten tonen dat wanneer de technologie slinkt, de schrijfvertraging toeneemt en zijn distributie wijder wordt door het effect van BTI.

ACKNOWLEDGEMENTS

As with every honorable life ventures, this Ph.D work profited from the efforts, wisdom, encouragement, unparalleled supervision and indispensable support and collaboration of a great number of people. Therefore, it is crucial to appreciate and acknowledge them. First and foremost, I would like to thank God the father, His son Jesus Christ and the Holy Spirit in a very special way, for divine directions, mercies, wisdom, strength, courage, protection, providence and guidance to me throughout my Ph.D pursuit.

I would like to say a very big thank you to my daily supervisor and promotor, Prof. dr. ir. Said Hamdioui for accepting and providing me with an opportunity to pursue my Ph.D thesis under his guidance. Said, thank you for an outstanding enthusiasm, availability, encouragement, mentorship, kind hearted, and supervision from the onset of this research which is instrumental in seeing me through the turbulent research days. I learnt not only how to research, clear and structural scientific paper writing, and good presentations but also the art of having strong opinions, dedicated and hardworking researcher via your exemplary leadership. Furthermore, you provided a good learning environment through collaborations with *IMEC*, participating in workshops, seminars, and summer schools. *Thank you once again!* In addition, I want to acknowledge Dr. ir. Mottaqiallah Taouil for his encouragements, assisting with *MATLAB* scripts, correcting my papers, creating a conducive working environments and being available for brain-storming sessions. *Thank you!* Furthermore, I want to thank Dr. Seyab Khan for my early discussions with you at the onset of this project. Moreover, I want to specially acknowledge the head of my department, Quantum and Computer Engineering, Prof. dr. ir. Koen Bertels. Thank you for the wisdom, care and efforts in creating a nice environment and atmosphere for research in both the group and the department. Furthermore, I would like to thank the rest of the committee members for assuming their roles, reading this dissertation, and delivering feedback. *Thank you!*

I want to specially thank Prof. dr. ir Francky Catthoor, Dr. Praveen Raghavan, Dr. ir. Halil K*ü*kner, Dr. Pieter Weckx, Dr. Stefan Cosemans, and Prof. dr. ir Wim Dehaene of *IMEC* and Katholieke Universiteit Leuven, ESAT, Belgium for a fruitful collaborations we had together in providing us with both model and designs used in our simulation experiments. In addition, I want to thank Francky for his ever readiness and willingness to read our manuscripts and give a prompt inputs that makes our work better *Thank you!* Moreover, I want to thank Halil and Pieter for the time they spent in putting me through the Atomistic *BTI* model.

I would like to express my appreciation to the Computer Engineering (CE) secretariat and staff for taking care of all the administrative matters on the course of my programme and for this, I want to say thank you to Lidwina. Eric and Eef, thank you for creating and

updating the websites with all the publications, managing the servers, resolving computer issues, and installation of several software and so on.

I would like to thank Imran for organizing the weekly CE football matches. These matches refreshed and gave me clear mind to think through the challenges that come with my research task. Thanks to everybody that took part in these matches. Moreover, I want to thank Miki, Anh, Joyce, and Carmina for organizing monthly Q&CE department colloquium. Thanks to all who attended and contributed to the insightful discussions.

I would like to also thank my previous and current office mates. Demid, Mahyar, Motta, Seyab, Mafalda, Mahroo, Lei, Anh, Peyman, Daniël, Lizhou, Muath, Jintao, and Guilherme. Thanks for all the lively and interesting discussions! I want to also thank Daniël for translating both my propositions and dissertation summary to Dutch. I want to also thank Guilherme for designing the cover of my dissertation. In addition, I want to acknowledge Prof. dr. Sorin Coţofană for my interactions with you which gives me useful experience and also prepared me for this Ph.D program. *Thank you!*

I want to thank in a special way all the members of Mount Zion Parish of the Redeemed Christian Church of God, Delft. Thanks so much for all the refreshing time of fellowship we shared together. Thanks to Pastors Hyacinth and Helena, Salifu, Jerry and Joy Okpanachi, Jide, Frank and Evelyn Annor, Afred and Glory Larm, Abimbola, Sam and Linda, Chidi, Theresa, Edna, and all the other church members.

I want to give special thanks to Pastors Walter and Betty Ononogbu, Dr. ir. Alex and Aderonke, ir. Pastor Tope and wife, Pastor Peter Okwudili Okekeorji, Jide Ogunbo, Dr. Pauline Ikpa and all the Predestined family, Mummy Oby Chukwu, Pastor Basil Nwatu, Engr. Gabriel and Nwuka Ugwuchukwu, Tpl. Alex and Chinwe Ogbodo, Engr. Chukwudi and Ijeoma Agbo, Tpl. BB and Ifeoma Ogbodo, Hon. Peter Nnaji, Mr. & Mrs. Ikechukwu Ngozi Okenwa for all your prayers and support. In addition, I am also indebted to my friends; Austine and Chinwe Ajah, Beckley and Matilda Nfor, Sandra, Sam and Chinelo Menkiti, Peter, Fredrick, Dayo and Bukayo, Gerard Fossung and family, Sophia, Sam and Obioma Ani, Ezra and Nneka Mallo, Christian, Rostand and Henrieke, Robert, Chibuike, Arc. Emile Jansen (my Dutch teacher) and Giovanni Anthony (my neighbour) for all their wonderful assistance.

Words are inadequate to express my heart-felt appreciation of the love, care and prayers of my lovely mother, Mrs Grace Ngozi Agbo, sisters, brothers and in-laws. They have greatly motivated me to continue. Lastly, I want to deeply thank my better half, Amarachukwu for her patience, understanding and prayers to me while undergoing this Ph.D work. *Thank you so much!* To all who have inspired and supported me in one way or the other, but could not be mentioned here due to space limitation, I sincerely thank you.

Agbo I.O.,
Delft, June, 2018

CONTENTS

Summary vii

Samenvatting ix

Acknowledgements xi

1 Introduction 1
- 1.1 Motivation . . . 2
  - 1.1.1 Transistor scaling . . . 2
  - 1.1.2 Impact of scaling on variability and reliability . . . 5
- 1.2 Classification of reliability failure mechanisms . . . 7
- 1.3 The state-of-the-art in memory reliability analysis . . . 14
- 1.4 The state-of-the-art in reliability mitigation schemes . . . 15
- 1.5 Challenges . . . 19
- 1.6 Research Topics . . . 20
- 1.7 Contributions . . . 21
  - 1.7.1 Analysis of aging impact on different memory sense amplifiers . . . 22
  - 1.7.2 Investigation of aging impact and mitigation on memory read-path . 23
  - 1.7.3 Investigation of write path aging . . . 23
- 1.8 Thesis Organization . . . 23

2 BTI Impact on different Memory Sense Amplifiers 25

3 BTI Impact and Mitigation on the Memory Read-Path 89

4 BTI Impact on the Memory Write-Path 103

5 Conclusion 111
- 5.1 Summary . . . 112
- 5.2 Future Research Directions . . . 113
  - 5.2.1 Memory Aging Modeling . . . 113
  - 5.2.2 Mitigation schemes . . . 115

References 117

List of Publications 133

Curriculum Vitæ 137

1

INTRODUCTION

1.1 MOTIVATION
1.2 CLASSIFICATION OF RELIABILITY FAILURE MECHANISMS
1.3 THE STATE-OF-THE-ART IN MEMORY RELIABILITY ANALYSIS
1.4 THE STATE-OF-THE-ART IN RELIABILITY MITIGATION SCHEMES
1.5 CHALLENGES
1.6 RESEARCH TOPICS
1.7 CONTRIBUTIONS
1.8 THESIS ORGANIZATION

In the recent years, transistor scaling has led to significant performance enhancement and reduced the cost of electronic systems. At the same time, it has introduced new reliability challenges in nano-scaled devices. Memories are generally constructed with the smallest node transistors to increase their size and capacity. Therefore, understanding the impact of reliability failure mechanism on such memories is crucial to provide insight to integrated circuits (IC) designers. In this chapter, we first provide the motivation for memory system reliability. Second, we present a classification of the reliability failure mechanisms. Third, we provide the state-of-the-art in memory reliability analysis. Fourth, we present the state-of-the-art in reliability mitigation schemes. Fifth, we present the main research challenges with respect to memory reliability. Sixth, we describe the research directions of this dissertation, as well as the main contributions. Finally, we provide the outline of the dissertation.

1

1.1. MOTIVATION

THE purpose of this section is to familiarize the reader with (memory) reliability in the nano scale era. Section 1.1.1 gives a brief background on transistor scaling. Section 1.1.2 explains how scaling affects the variability and reliability of a transistor.

1.1.1. TRANSISTOR SCALING

The benefits of CMOS technology scaling are evidently visible in every aspect of our daily lives. Technology scaling has led to faster, denser, and cheaper electronic products. This is due to high performance, smart design architecture, and high transistor integration per unit area known as integrated circuits (IC). Large ICs are equally known as Very Large Scale Integration (VLSI) circuits. The VLSI design has been around for over four decades, also starting with the Intel 4004 microprocessor, which incorporated 2000 transistors per unit area on a die of 12 mm^2 [1].

According to Moore's law, the IC's density doubles nearly every two years [1]. Moore's law has been regarded as a road-map to the whole semiconductor industry. Again, upholding the Moore's law has led to constant technological innovation. Although, technology scaling is still maintaining its benefits in terms of IC density, the benefits in terms of performance and cost already started declining [2]. Technology (or transistor) scaling can be typically categorized into conventional and innovative approaches; they are explained next.

CONVENTIONAL SCALING

Traditional scaling can be classified into two categories:

- Constant voltage scaling: It refers to the case where the transistor dimensions decrease per each technology node at a fixed supply voltage.
- Constant field scaling: It refers to the case where both the transistor dimensions

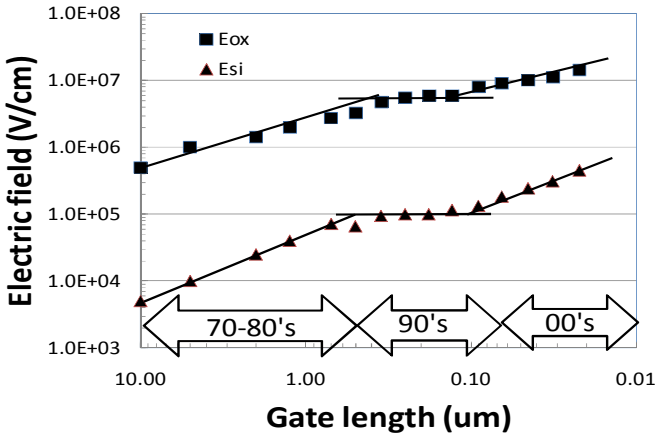


Figure 1.1: Evolution of oxide and silicon electric fields [2,3].

and supply voltage decrease simultaneously as the technology scales down.

The Figure 1.1 [2,3] depicts the history of traditional scaling. The y-axis gives the electric field which denotes both the oxide (*Eox*) and the silicon (*Esi*) fields, while the x-axis shows the three phases, the scaling experienced.

- The first phase is between 1970 and 1980 which is based on constant voltage scaling. It has caused an increase in the electric field which implies more stress (that is severe reliability).
- The second phase is the 1990's which is based on constant field scaling. It results into a constant electric field indicating that stress is not becoming worst (that is stable reliability).
- The third phase started in 2000; it is based on constant voltage scaling. The supply voltage cannot scale anymore for different challenges such as: leakage and performance. Clearly, the third phase has led to an increase in electric field which may result in the acceleration of the transistor degradation. This has ended what is called Dennard scaling [4,5] around 2005 and 2007, meaning a constant improvement factor on the transistor's performance, power, and density needs. Dennard scaling has stopped beyond 45 nm technology node due to increase in leakage current. Only Moore's cost scaling has largely continued up till now.

The transistor area scaling is continuous while the performance scaling requires an enhancement techniques to keep maintaining the Moore's law scaling road-map. These performance enhancement techniques called innovative scaling will be described next.

INNOVATIVE SCALING

In addition to the scaling of device dimensions, enhancement techniques have been developed not only to keep upholding the Moore's law scaling phenomenon but also to further increase the overall performance, reduce the leakage, etc. There are mainly three performance enhancement techniques and they are briefly explained next [6,7] in Figure 1.2:

- **Channel strain engineering:** This is one of the innovative techniques to maintain the unending scaling of CMOS technology to nanoscale nodes such as 90 and 65 nm. There are two types of channel strain engineering technologies: 1- global, and 2- local. Global strain engineering is the application of stress (thin strained Si layer) on the whole substrate (SiGe) before manufacturing the device, while local strain engineering uses one of these techniques (such as shallow-trench-isolation, epitaxial layers, and / or highly stressed nitride capping layers) to cause stress (thin strained Si layer) in either NMOS or PMOS device. The difference between global and local strain engineering is that the former produces larger strain than the local type.

  Local channel strain technique has more advantages than the global ones; hence, more details of latter will be given in this work [8]. Local channel strain engineering is proposed first to overcome the speedy rise in gate tunnelling current which is as

a result of decrease in both gate oxide and supply voltage as the CMOS technology scales down. Second, the channel strain is proposed to maintain the performance high with technology scaling. This technique aims at increasing the movement of both electrons and holes in the channel while maintaining an approximate oxide thickness. The movement of holes in the channel is known as compressive while that of electrons is known as tensile. The compressive develops SiGe material on the PMOS source-drain device while tensile develops SiN material caps on the NMOS device [7].

- **High K / Replacement Metal Gate:** The prior strain technique has successfully sustained the CMOS technology scaling enhanced performance for two technology nodes. This technique failed to sustain the scaling any further. This is because the SiO_2 CMOS device gate insulator has attained its physical limits. As a result, scaling the device oxide thickness beyond 65 nm was no more possible leading to high gate leakage. High-k/metal gate technology as was invented as the solution to the high gate leakage that is to keep CMOS scaling at a lower gate leakage [9,10]. The high-k/metal gate technology changes the classical SiO_2 dielectric with Hafnium-dependent high-k gate dielectric layer [7].

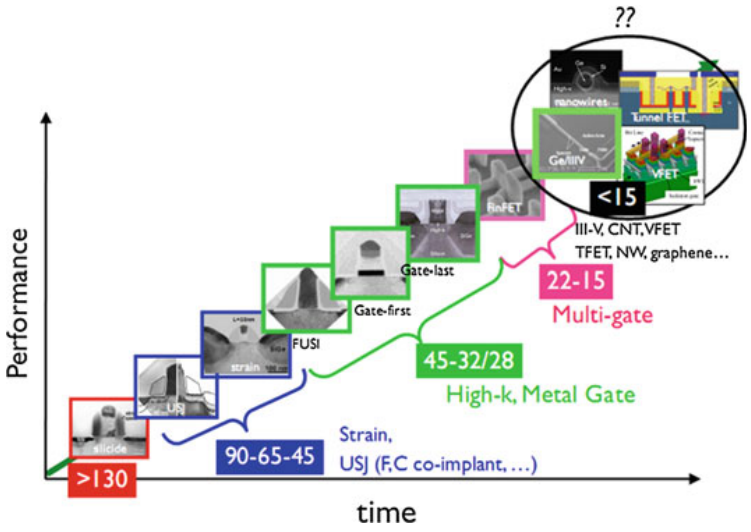


Figure 1.2: The roadmap for CMOS technology. A lot of newer device options are currently being investigated for future nodes [7].

- **Tri-gate / FinFET:** The continuous CMOS scaling of the device dimensions beyond 32 nm technology has faced with the Short Channel Effects (SCE) issue. SCE is the alterations of the transistor's threshold voltage with the reduction of the device channel length which is observed in deep sub-micron CMOS technology node. An

example of short channel effect is Drain-Induced Barrier Lowering (DIBL). DIBL occurs when the potential barrier is reduced by increasing the drain voltage so as to allow the flow of electron between the source and the drain of the device. The SCE has caused a serious gate channel length wall which also lead to the device threshold voltage wall.

However, to resolve this problem, Tri-gate/FinFET technology has been proposed. The FinFET and Tri-gate technology aims at extending the control of the gate on both sides for FinFET and three sides for the tri-gate [11,12]. This is realized by changing the CMOS Planar 22 nm with a 3D tri-gate architecture. The 3D architecture is made up of three sides for both tall and narrow silicon fin. These sides are covered by the gate stack. A decreased OFF-state leakage and increased performance is caused by fully depleted operation which also improved electrostatic channel control of the gate. Moreover, optimizing the channel control can lead to further down scaling of the gate length. Therefore, to obtain more performance benefits, a hybrid technique is required; it integrates a new 3D device technology, strained silicon and high-k/metal gate solutions [7] as shown in Figure 1.2. In addition, so called boosters or Design-Technology Co-Optimisation (DTCO) techniques have been emerging and added on top of all these for 14-10 nm mode [13].

1.1.2. IMPACT OF SCALING ON VARIABILITY AND RELIABILITY

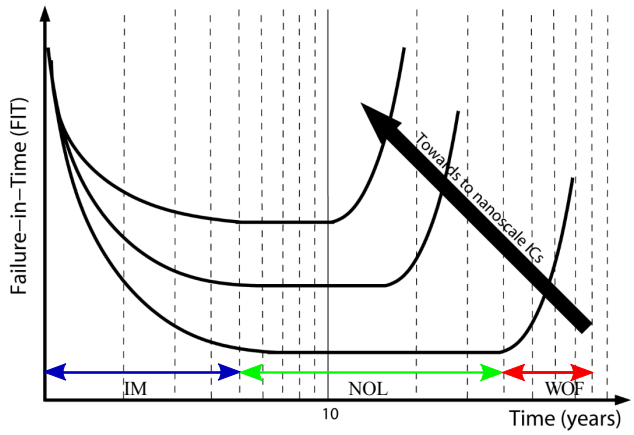


Figure 1.3: Bathtub curve depicting the generic shape of the failure rate of a chip comprised of n statistical individual parts (tailored from NASA, 2008) [14,15].

From the technology point of view, reliability is now more than ever a major bottleneck with the down-scaling continuation of CMOS technologies. This is due to increase in electric fields and power densities attaining the values that can be permitted for reliable operation. Moreover, the technology scaling impact on reliability can be considered with respect to failure rate as a function of time which is illustrated using a bathtub curve, Figure 1.3. The figure depicts the failure rate of ICs which are produced in three different

technology nodes. The y-axis represents the failure rate while the x-axis represents the operational life time of the chips. Each bathtub curve is divided into three phases: infant mortality rate (IM), normal operating life (NOL), and wear out failure (WOF). These three phases are explained next.

INFANT MORTALITY RATE

This is the initial stage of the operational lifetime of the fabricated chip. At the fabrication stage, circuits or chips are subjected to a severe conditions and at high endurance figures in which the chip may stop functioning as desired at their initial operational lifetime. As a result, the failure rate in this phase is high, indicating big possibility of chips or circuits that might stop working as they do not meet the expected requirements. Chips which cannot survive at a very severe circumstances will be detected, removed, and set aside [14].

NORMAL OPERATING LIFE

Normal operational lifetime of the chip is the second stage after the chip has been fabricated. The chip or circuit functions according to the proposed descriptions at the design time. Note that the failure rate is expected to be very low at the normal operational lifetime. The failures may be caused by variations, radiation, and overshooting the permitted operational state, etc [14].

WEAR OUT FAILURE

The wear out failure is the third stage of a chip's lifetime after fabrication. There is an increment in aging speed of the chip when the latter starts to wear out or get exhausted. Wear out failure mechanisms are caused by soft errors, electron-migration, Bias temperature instability, time dependent dielectric breakdown, hot carrier injections and so on [14]. It is worth noting that some of the causes of the wear out will be briefly described later in this dissertation. In order to extend the normal operational lifetime of the chip; the chip designer or fabricators are expected to take this wear out stage into consideration and employ safety margins and/or mitigation techniques in advance (at the design stage). This is because, once the chips are delivered to the customers, the chip designers do not have any regulation over the chips any longer [14].

Taking a look at Figure 1.3 reveals the following impact of technology scaling:

- **Higher failure rate during operational life:** The figure depicts that as the CMOS technology scales down to the nanometer regime, the failure-in-time increases (moves upwards) with respect to the first bathtub curve closest to the x-axis (time); this will lead to serious device unreliability and therefore requires an urgent actions either at design time or during the operational lifetime. Nevertheless, as recent work shows that the aging is strongly workload-dependent, variable failure rate in this phase is expected, with most failures being transient and not fatal any longer! Hence, the total life time due to fatal functional device failures has not decreased really. It is the parametric and transient errors which cause the practical problem.

- **Reduced lifetime:** The figure also depicts that as the technology scales down to the nanometer regime; the lifetime of the chips reduce (up to half that is tending to the left); this implies increase in the cost of purchasing the same electronic system within a short period of time. Therefore, there is need to address this problem and elongate the lifetime of the chips either at the design time or at the operational lifetime.

The above are major reliability concerns. Critical and long lifetime application require exactly the opposite: lower failure rate and long lifetime.

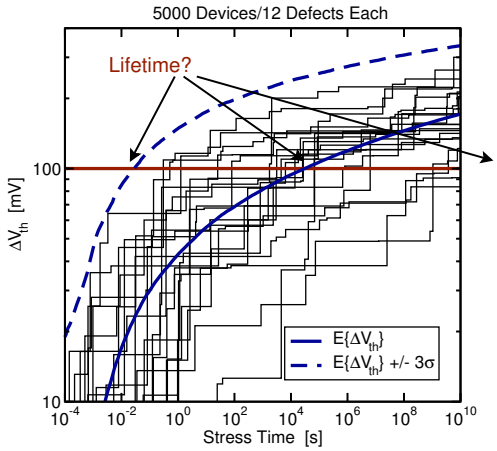


Figure 1.4: Life time in the presence of variability [16,17].

It is worth noting, that Figure 1.3 does not hold any more for deeply scaled technology nodes [18]. Figure 1.4 shows how the Vth distribution evolves over time for 5000 measured devices each while fixing the accurate margin of Vth life time for the devices at 100 mV. The figure also shows that the mean Vth for time dependent variability of the devices already failed after 10^4s before reaching its expected lifetime. Moreover, the figure also shows that when the variation of +/- 3σ is added to the mean variations, some of the devices started failing even earlier. All of these indicate that the traditional bathtub curve cannot hold any more in the presence of time-dependent variability.

1.2. CLASSIFICATION OF RELIABILITY FAILURE MECHANISMS

Figure 1.5 depicts the two classes of reliability failure mechanisms: time zero and time dependent mechanisms. Next, these will be briefly described.

SOURCES OF TIME ZERO VARIABILITY

Time zero (t = 0) variation occurs at the start of the chip's lifetime due to imperfect device fabrication process steps; for example, polishing, lithography, resist, etching, doping

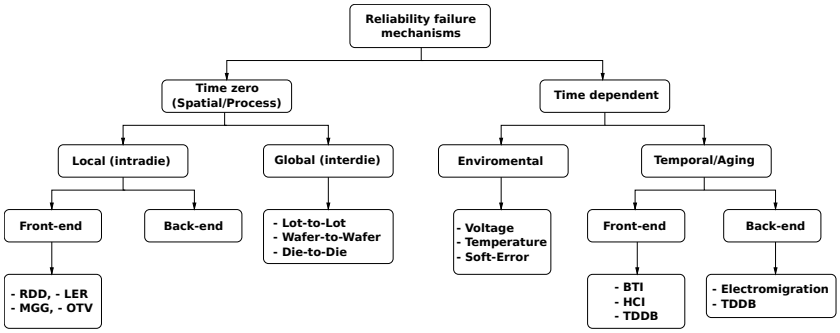


Figure 1.5: Reliability failure mechanisms classification [6].

and so on [19,20]. The device parameters are deviated as compared to the expected ones at the manufacturing phase due to the limited fabrication process controllability [21]. There are two types of time zero variability: 1 - Local variation and 2 - Global variation [22].

LOCAL VARIATIONS

This variation is due to parameter mismatches across the same manufactured devices or interconnects for a short distance inside a die [23–25]; for example variations in length, width, oxide thickness, flat band control, number of dopants etc. [26–28].

Figure 1.5 depicts the local variations divided into front-End-Of-the Line (FEOL) and Back-End-Of-the Line (BEOL). The FEOL comprises of crucial intrinsic parameter fluctuations such as Random Discreet Dopants (RDD) [19,23,29–39], Line Edge Roughness (LER) [19,35,37,40–45], Metal Gate Granularity (MGG) [19,30,37,46–49], and Oxide Thickness Variations (OTV) [23,29,30,34,50–52]; these intrinsic parameter fluctuations are briefly described next.

Random Discreet Dopants (RDD): Obviously, RDD causes intrinsic parameter variations of field effect MOS transistors [38,39]. RDD is due to the unending down-scaling of CMOS technology [19,29]. This scaling reduces the transistor dimensions which directly reduces the number of dopant atoms per transistor [36]. Most of the transistor parameters such as threshold voltage V_{th}, Sub-threshold Slope (SS), cut-off frequency (F_t), gate capacitance (C_g) and so on, are being affected as a result of random position of the atoms [23]. And the channel doping of the MOS transistors affect these parameters. Moreover, the random dopant mostly affects V_{th} parameter of the MOS transistors. Saha in [23] gives an expression for the number of dopant atoms which is stated below:

$$N_{total,chan} \cong N_{CH}[W_{eff} L_{eff} \chi_j] \tag{1.1}$$

in which W_{eff}, L_{eff}, N_{CH}, and χ_j denote the effective channel width, effective channel length, the doping concentration of the channel, and the extension junction depth of the source-drain, respectively. Equation 1.1 shows that the number of dopant atoms shrink due to dimensional scaling.

Line Edge Roughness (LER): LER is another source of threshold voltage variability; and it is observed at time zero variability. The gate pattern edges are irregularly fluctuated which is known as LER [53]. The sub-wavelength lithography and etching processes are the main causes of LER [19,35,37,43]. Patterning of transistors are realized by using sub-wavelength lithography for earlier technology node such as 250 nm node [40–42,44,45]. Moreover, LER was not a challenge for 250 nm and higher nodes while it could be a challenge for 22 nm and lower nodes. LER did not scale in the same manner as CMOS technology and it is not reliant on the applied lithography type. LER also hinges on the transistor gate length and as such it scales as CMOS technology. Therefore, the gate length scaling is a vital cause of V_{th} variability while leading to leakage current increment. Consequently, with the unending CMOS technology scaling into deep nano-era regime, the LER is foreseen to be the next candidate to succeed the random discreet dopant as a latest and vital of fluctuation [41].

Metal Gate Granularity (MGG): The high-K materials poses two main challenges to the poly silicon gate. These challenges are Femi level pinning and phonon diffusion. These challenges are not adapted with the poly silicon gate. They cause both increase in threshold voltage and decrease in the movement of the electron. However, to overcome these issues, the poly silicon gate is succeeded by the metal gate granularity; hence, the integration of high-K and metal gate materials known as high-K / metal gate. This is used in 45 nm technology node [19,30,37,46,48]. The metal gate granularity brings reduced gate resistance while leading to a rise in ON current. Polycrystalline is a type of metal gate which is used in this analysis. It consists of different sizes of grains, orientation based on the manufacturing circumstances [46]. The metal crystallization is caused by the after effect of metallic coating recovery which is an issue for the initial gate technology [47]. Moreover, the random fluctuations of the transistor gate threshold voltage is caused by varying work functions at the interface of the metal / high-K material in the metal grains of various crystal arrangements [48,49]. Therefore, gate last technology can be used to avoid the variations caused by the metal gate granularity that is in the polycrystalline gate material. Overall, the threshold voltage variability in the metal gate granularity hinges on the size, work function and the arrangement in the grain [46].

Oxide Thickness Variation (OTV): OTV is caused by continuous aggressive scaling of the CMOS technology into the sub-micrometer regime. The OTV causes the threshold voltage variability. The OTV also originates from the atomic scale roughness at the Si/SiO2 interface of the transistor gate area [29,30,34,52]. Moreover, the main oxide thickness variation at the gate of either NMOS or PMOS device is caused by the time the thickness of the oxide is equal to a handful of silicon atomic layer; as a result leading to the atomic scale interface roughness steps [23,29,50]. Therefore, the sub-micrometer MOS devices differ from each other as result of variations in both oxide thickness and its interface causing main fluctuations for threshold voltage while less effect for mobility and gate tunneling current. Subsequently, if the oxide interface length is equivalent to the transistor sizes, the variation due to the oxide thickness will be much more [50]. Empirically, it has been demonstrated that both oxide thickness and doping variations have

an equivalent effect on threshold voltage distributions that is for planar technology at sub-30 nm transistor gate length [29,50,51]. With more complex gate stacks this list has further increased for the most recent nodes but that is not the focus here.

GLOBAL VARIATIONS

The parameter variation between dies which emerge from various runs, lots, and wafers are known as interdie (global) variations. This variation is caused by the parameter alterations on the same device/interconnect over larger area or manufactured at varying time, and is a result of factors, such as processing temperature, and equipment/tool properties, etc., between various runs, lots, wafers and dies. It leads to five different process corners: Typical-Typical (*TT*), Fast-Fast (*FF*), Slow-Slow (*SS*), Slow-Fast (*SF*) and Fast-Slow (*FS*) [54]. These process corners are grouped into two classes and they are stated as:

- **Uniform corners:** The uniform process corners include (*TT*), (*FF*), and (*SS*). This is because they impact the NMOS and PMOS devices uniformly.
- **Non-uniform corners:** The non-uniform process corners include (*SF*), and (*FS*). They affect the NMOS and PMOS devices unevenly and affect their time zero characteristics.

In addition to time zero variation, there is another type of variation at time greater than zero and it is known as time dependent variation. This type of variation will be explained next.

SOURCES OF TIME DEPENDENT VARIABILITY

Time dependent (t > 0) variability is more pronounced and appreciated than ever before due to technology scaling in the decananometer regime [55]. There are two sources of time dependent variability; they are environmental and temporal (aging) variations. This section will briefly discuss the sources of environmental variations and afterward, the most vital aging failure mechanisms as depicted in Figure 1.5.

ENVIRONMENTAL VARIATIONS

Environmental variations are changes which occur during the operation of a circuit. Therefore, it is crucial to ensure that a circuit meets its requirements at all time to prevent time dependent environmental variations. We focus on the impact of temperature, supply voltage, and soft error in this work.

Supply voltage: Supply voltage variations impact the operating speed of MOS transistors. The fluctuation in switching activity across the die / circuitry causes a non-uniform power / current requirement and may lead to logic failures. In addition, transistor subthreshold leakage variations impact the non-uniform distribution of supply voltage across the circuitry as well. Therefore, the reduction in supply voltage degrades the performance of the circuit / transistors and increasing supply voltage compensates / improves the performance and significantly lessens circuit failure rates due to variability [56].

Temperature: Temperature variations impact the operating condition of MOS transistors. Hence, temperature rise reduces the threshold voltage (which positively impacts the delay); and decreases the carrier mobility (which negatively effects the delay) and consequently increases the leakage current [57]. The dependence of threshold voltage and temperature is given by [57].

$$V_{th} = C_{vt} - \frac{Q_{ss}}{C_o} + \Phi_{ms} \tag{1.2}$$

where C_{vt} denotes a constant that represents fermi potential, surface charge Q_{ss} at the Si-SiO_2 interface, gate oxide capacitance C_o and work function difference Φ_{ms} represents a function of the temperature [57].

$$\Phi_{ms} = -0.61 - \Phi_F(T) \tag{1.3}$$

Here Φ_F(T) denotes Fermi potential. Expression 1.3 depicts that work function difference reduces with respect to increase in temperature and hence decreases the threshold voltage.

Soft-Error: Nowadays, the sudden rise in the radiation susceptibility is due to the increasing quest for lesser power and denser transistors on a chip; this is to keep up with aggressive reduction in the sizes of transistor dimensions and its operational voltage as predicted by Moore's law [1]. Moreover, the reduction in the transistor dimensions beyond 1 μm of the channel length can severely affect the functionality of the storage devices. The effects are short-circuiting, circuit disturbance, and off-state terminal draining, which are due to the cosmic rays striking of the storage devices [58]. These cosmic rays were first observed in space borne electronics and similarly on Dynamic Random Access Memories (DRAM) [58]. In the case of memories, errors manifest themselves when a data is written into the memory cell which has been arbitrarily altered while the memory cell is not disfigured. This is known as a soft error; and the occurrence rate of this error is known as soft error rate [59]. The mathematical model to determine the soft error rate is given by [60,61]:

$$SER \propto F \cdot A \cdot e^{\frac{Q_s}{Q_{crt}}} \tag{1.4}$$

In Equation 1.4, F denotes neutron flux, A denotes the circuit area affected by the striking particle, Q_{crt} denotes the critical charge, and Q_s denotes the transistor accumulated charge efficiency. The angle and energy of the striking particle on the transistors of memory circuit are the drawbacks of this model which is not taken into account while developing the model.

TEMPORAL/AGING VARIATIONS

TEMPORAL variability is a failure mechanisms (i.e., aging) that slowly widens the initial distribution of the IC based on the usage of the system and its environmental conditions [62]. There are four kinds of reliability failure mechanisms. They are Bias Temperature Instability (BTI), Hot-Carrier Injection (HCI), Time Dependent Dielectric

Breakdown (TDDB), and Electro-migration (EM); though BTI is considered the most important failure mechanism in the research community and it will be the main focus in this work. Next, the four failure mechanisms will be briefly explained.

Bias Temperature Instability (BTI): BTI is widely claimed and researched aging phenomenon in the reliability community [63–66]. It is a reliability challenge in today's nano-scaled technology [67–69]. BTI is observed in MOS transistors due to an increment in absolute threshold voltage V_{th}, decreased drain current at a high temperature and for a biased transistor gate voltage over their operational lifetime [67,70–73]. BTI is classified into two types and they are Negative BTI (NBTI) and Positive BTI (PBTI). NBTI affects the PMOS type transistor and PBTI affects the NMOS type transistor. The V_{th} decreases in PMOS device when NBTI is applied while V_{th} increases in NMOS device when PBTI is applied.

There are two well-known BTI models and they are Reaction Diffusion (RD) model and Atomistic trap based model. RD model is one of earlier developed models for analyzing the impact of aging in MOS devices [74–78]. RD is based on the concept of breaking of silicon hydrogen bond at the silicon dioxide SiO_2 interface of NMOS or PMOS transistor [63]. As a result, RD model has two phases consisting of stress phase and the relaxation phase. At the stress phase, the silicon hydrogen bond is broken at the gate of either NMOS or PMOS device. The broken silicon bond stays at the interface (traps or charges) while either the hydrogen or its molecule diffuses towards the device gate. However, at the relaxation phase, there is no breaking of silicon hydrogen bond; as a result, the hydrogen or its molecule diffuses back to the interface which is also known as recovery or annealing.

Though, RD model has been mostly used by a number of researchers but there are still questions whether it is accurate enough as an ideal or near ideal aging model. Notwithstanding, all the questions posed on the RD model, it is still been used by researchers in their analysis work. Nevertheless, Atomistic trap based model was developed as alternative and accurate aging model. Atomistic model is based on the assumption that during the manufacturing of the chip, the chip has a latent defects on it which is only activated when the chip is stressed or charged, resulting into an activated traps. These traps are function of temperature, supply voltage, duty factor, and workloads [67,73,74,79–92].

It is worth noting that according to literature, the PBTI did not occur until CMOS technology scales down to 45 nm node [28]. This implies that NBTI has been the aging phenomenon investigated for higher technologies than the 45 nm [93–96]. Though, in literature some researchers have investigated the impact of PBTI for higher technologies such as 90 and 65 nm. The researchers argue that the introduction of high-K / metal gate material for the 45 nm so as to keep up with CMOS scaling results in PBTI effect NMOS transistor [97,98]. The latter shows that PBTI (NMOS) causes charge defect in high-K / metal gate while NBTI (PMOS) leads to interface traps at silicon dioxide channel gate. Overall, the BTI is strongly dependent on the workload (application) stress patterns [99–101].

Hot Carrier Injection (HCI): Hot charges inside the NMOS and PMOS transistors cause a degradation phenomenon known as HCI [102,103]. These hot charges are due to the

movement of current from the source to the drain of the transistor across the channel at an excited energy level as compared to the mesh temperature. As a result, sufficient energy level is gained and introduced at the transistor's gate oxide; the charge traps are obtained at the interface state. The performance parameters of the device are affected due to the interface state. On top of that, hot charge injections are caused during the switching states at time of its operation. In addition, there is a proportionate relationship between the key device parameters (i.e., channel length, Tox, and V_{dd}) and the rate of introducing hot charges inside the transistor gate oxide [104–107]. Overall, simultaneous decreasing of V_{dd} and transistor dimensions are not at same rate with decreasing the length of the channel to achieve excellent performance. Therefore, the higher the current densities, the higher the sensitivity of the transistor to hot charges.

Time-Dependent Dielectric Breakdown (TDDB): TDDB is another wear-out technique which can be originated in two steps i.e., Front-End-Of-the-Line (FEOL) and Back-End-Of-the-Line (BEOL). It also degrades both devices and interconnects. There are two types of TDDB, being high-K TDDB and low-K TDDB. The high-K TDDB affects the device reliability while low-K TDDB affects the on-chip interconnect reliability. They are briefly described next.

- **High-K TDDB:** This is another crucial source of failure mechanism which endangers the gate reliability [107–110]. There are various mechanisms that are proposed, e.g., charge injection, interface, bulk trap state generation, and trap assisted conduction [106,108,111,112]. Though, there is no accord reached in the research community on a specific one yet.

  Several research works [106–108], have reviewed the TDDB failure phenomenon. Their reviews reveal that applying an electric field over the device gate leads to constant deterioration of the dielectric material. Therefore, it leads to the creation of a conducting paths in oxide that inhibits the anode and the cathode. The occurrence of the latter leads to an extreme loss of energy inside the breakdown path due to relentless effect of electric field over the gate oxide. As a result, the breakdown causes two significant overheads such as 1- loss of control of the flow of current at the transistor gate and 2- rapid increment in the current of the gate. The unending down-scaling of CMOS technology leads to decrement in gate-oxide thickness while increment in the gate-dielectric leakage current and TDDB impacts.

- **Low-K TDDB:** The low-K TDDB is another failure mechanism event for the interconnect or Back-End-of-Line (BEOL) reported in literature by researchers as the CMOS technology scales down to the deca-nanometer regime [113,114]. Though, the integration of low-K material and Cu interconnect has been accepted because of lesser resistance-capacitance delay, crosstalk noise, etc., for the 90 nm node. But the SiO_2 dielectric breakdown strength between the wire is stronger than that of the Low-K material [115,116]. The breaking of the weak dielectric bond of low-K material is supposed to cause the TDDB failure [117–119]. This failure leads to a sudden increase in leakage current. The modeling of low-K TDDB failure mechanism is a function of the electric field. This electric field is a function of time which causes device mismatch, thereby affecting its lifetime [116].

Electro-Migration (EM): Electro-migration is another type of failure mechanism that has been studied in literature. This is the change in kinetic energy at both electrons and metal atoms which cause the motion of the ions within the conducting element and this leads to an increase in current density. This is due to continuous reduction of metal wires with a corresponding rise in the wire resistance [28,120–124]. The interaction between electrons and metal atoms results to voids and hillocks [121,124].

The electro-migration can be illustrated for a temperature, T > 0K which induces the atomic oscillations. The spreading of electrons and eventually the movement of electrons in a specific directions is due to the atomic oscillations. Furthermore, the application of velocity on a mass of an electron and metal atom (known as transference of momentum) affects the electro-migration. The high doping of silicon causes the silicon to behave as a metal known as a polycrystalline silicon. The creation of a void as a result of atomic oscillation, enables the movement of an atom to fill that void.

Overall, the EM phenomenon for high current densities can result to the metal failure, irrespective of the metal in question [125–127]. Therefore, the EM can be compensated by using a laminating barrier such as tungsten (W) and titanium(Ti). Again, Blech length limitations [128,129] can also be used to circumvent the electro-migration.

In addition, there are other techniques for analyzing and modeling which are beyond the scope of this dissertation [130] such as injecting (black-box controllability), detecting (black-box observability), and modeling reliability metric for a digital systems.

1.3. THE STATE-OF-THE-ART IN MEMORY RELIABILITY ANALYSIS

In recent years, many researchers have analyzed the reliability of the Static Random Access Memory (SRAM) systems. However, the focus has been mainly on the memory cell arrays, which typically consumes > 70% of a memory chip. The work on analyzing the impact of aging on the memory peripheral circuit (e.g., sense amplifier) is stilled limited; not to mention considering the whole memory system and the interaction between its different components. Next, a brief overview on the state-of-the-art on memory reliability modeling is given by dividing it in three classes:

- **Memory cell array:** Several works have studied the impact of aging on the SRAM cell arrays. For example, Kumar *et al.* investigated in [131] the impact of BTI on both Static Noise Margin (SNM) degradation and read stability of the memory cell arrays. Kang *et al.* investigated in [132] the effect of aging on the 6T SRAM cell design while exploring different performance parameters such as SNM, read and write failure probability, parametric yield of large SRAM array under process variation and memory leakage current (i.e., an I_{DDQ} measure). Krishnappa *et al.* compared in [133] various SRAM designs with regard to their reliability against BTI on the write margin, access time and leakage power metrics. Binjie *et al.* investigated in [134] the impact of aging on both Static Noise Margin (SNM) and Write Noise Margin (WNM) degradation of 6T SRAM cell. Bansal *et al.* studied in [135] the stability of an SRAM cell under the worst-case conditions and analyzed the effect of NBTI and PBTI, individually and in combination. Lin *et al.* compared in [136,137]

the impact of NBTI on SRAM V_{ccmin} stability and the combined impact of PBTI and NBTI V_t drift on the time dependent stability of SRAM V_{ccmin}, fabricated with high-K gate dielectrics. Weckx *et al.* investigated in [138] a combined statistical silicon which is extracted from large transistor arrays (32K) designed and fabricated in an advanced 20 nm High-K/Metal Gate process. In combination with current state-of-the-art based on statistical assessment approaches; these approaches acquire a realistic impact of BTI degradation on the yield and performance of 6T SRAM cells. Weckx *et al.* described in [138] the implications of BTI related time-dependent threshold voltage distributions on the performance and yield of SRAM cells. Hu *et al.* investigated in [139] the integral impact of both process variation and temporal degradation for FinFET technology while taking into account various memory cell arrays. Khan *et al.* analyzed in [140] a comprehensive aging degradation based on FinFET memory cell while considering various supply voltages, cell strength, designs, and technologies.

- **Memory peripheral circuits:** Few works have analyzed the impact of aging on SRAM peripheral circuits. For example, Khan *et al.* investigated in [141] the impact of partial opens and BTI in SRAM address decoder. Furthermore, Menchaca *et al.* investigated in [142] the impact of BTI on various sense amplifier designs carried out on 32 nm technology node by utilizing failure probability (i.e., flipping a wrong value) as a reliability metric.

- **Entire memory system:** Very few work has been done for the aging analysis of entire memory systems and the interactions between its components. For example, Kinseher *et al.* investigated in [143] the extent of integral effect of both degradation phenomena (i.e., BTI and HCI) for peripheral (particularly read and write circuits) SRAM design while considering different performance metrics of an industrial-based memory library. Kraak *et al.* analyzed in [144] the impact of BTI on high performance 14 nm FinFET memory design. It analyzes both the impact on individual components as well as the way they interact with each other while taking into account different workloads, the overall metric of the memory and also the individual components metrics.

Clearly, a lot of research efforts have been done on memory cell arrays while very few work has been done on memory peripheral circuits and one research effort on the entire memory system. There are many open issues to be still worked out e.g: Understanding and investigating the impact of aging on memory peripheral circuits, understanding and quantifying the aging rate of each memory part and how they interact with each other, etc. Note that various memory parts may degrade with varying rates depending on the workloads (applications).

1.4. THE STATE-OF-THE-ART IN RELIABILITY MITIGATION SCHEMES

Several research efforts are made to re-mitigate the impact of aging. The state-of-the-art in aging adaptation and mitigation techniques can be classified into four types

as illustrated by using Figure 1.6. They are worst-case design, design time aging-aware balancing, dynamic adaptation techniques, and adaptive resource management; they are explained next.

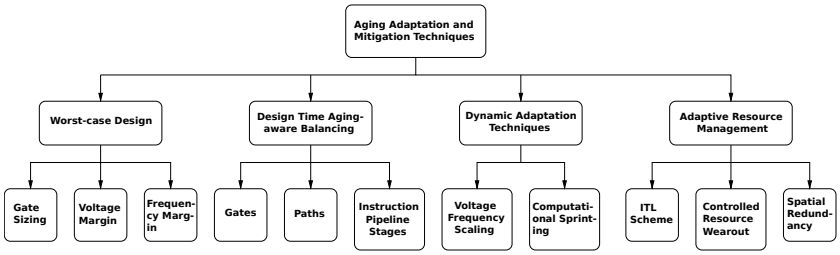


Figure 1.6: Classification of reliability mitigation schemes [145].

WORST-CASE DESIGN

Mitigation for worst-case design can be divided into three techniques, which are gate sizing, voltage margin, and frequency margin. The research efforts in these areas are provided next.

- **Gate sizing:** By changing transistor sizes, a more robust circuit against aging can be obtained. Much work has been done in this area. For example, Vattikonda *et al.* developed in [146] a predictive model for NBTI both for static and dynamic operations, while providing effective techniques to mitigate NBTI by using PMOS sizing. Yang *et al.* investigated in [147] a design approach to tolerate the aging degradation by gate sizing. They formulated an aging aware sizing problem and proposed a mitigation solution. Chen *et al.* proposed in [148] a new timing analysis flow, which can quickly and accurately predict path and gate degradation. In addition, they developed a new algorithm that can effectively identify the smallest set of critical reliability gates. This allows effective resizing with minimum area overhead. Kothawade *et al.* investigated in [149] a new technique for NBTI stress and mitigation at the instruction granularity. Their proposed NBTI mitigation technique modifies the physical register file by creating register banks and by transistors using sizing. Khan *et al.* proposed in [150] an NBTI gate delay model and a transistor sizing approach that mitigates the impact of NBTI based on path delay's. Kang *et al.* proposed in [151–153] an efficient transistor-level sizing algorithm based on an altered lagrangian relaxation approach to account for the temporal circuit degradation and to guarantee the lifetime of a circuit under NBTI.

- **Voltage margin:** By adding a margin to the supply voltage, aging can be compensated as the circuit runs faster at higher Vdd. For example, Zhang *et al.* proposed in [154] a prefixed/dynamic voltage schedule that gradually increases the operating voltage of the IC in order to mitigate the NBTI-related performance degradation.

- **Frequency margin:** By adding a margin to the frequency, aging can be mitigated as the delay of degraded paths will still fall within the clock period. For example, Lai *et al.* investigated in [155] the impact of aging on the clock distribution network and clock skew for various clock gating schemes. The cross-layer techniques is used to reduce the impact of aging on clock skew. Moreover, two Integrated Clock Gating (*ICG*) cell designs are used to change the idle state between logic high and logic low for each clock gating operation. Finally, a skew mitigation methodology was used to select the suitable *ICG* cells based on the architecture and micro-architecture context.

DESIGN TIME AGING-AWARE BALANCING

The delay of various paths of a system are typically balanced at time-zero. However, they could become imbalanced during operational lifetime as aging impact paths differently. By balancing the path delays for aging, a lower overall impact can be realized. It has been shown that the aging aware balancing can be carried out at three stages; they are at the gate, path, and instruction pipeline stage level.

- **Gate:** At this level, the mitigation aims at compensating an imbalance in rise and fall delays of gates. For example, Kiamehr *et al.* studied in [156,157] the impact of BTI and proposed a mitigation technique based on balancing the rise and fall delays of standard cells for the excepted lifetime.

- **Path:** At this level, the mitigation scheme tries to balance all paths within a pipeline stage. For example, Ebrahimi *et al.* proposed in [158] an aging-aware synthesis technique to increase the circuit lifetime by re-balancing paths with a specific guard-band.

- **Instruction pipeline stage:** At this level, the mitigation compensates for aging by balancing the pipeline stages. Oboril *et al.* proposed in [159] balanced pipeline design in terms of mean-time-to-failure, in which the pipeline stage delays are balanced for the desired lifetime. Their main idea is to create pipeline stages that are balanced for mean-time-to-failure rather than the traditional pipelines balanced for delay.

DYNAMIC ADAPTATION TECHNIQUES

The main challenge of the static mitigation techniques (i.e., worst case design, and design time aging-aware balancing) is that they need an accurate prediction of the aging degradation. Hence, over-estimation is typically required to tolerate the lifetime of the circuit. As a result, dynamic adaptation schemes are investigated. They can be divided into two types: 1- voltage and frequency scaling, and 2- computational sprinting.

- **Voltage and frequency scaling:** To reduce the cost of a fixed voltage and/or frequency margin scheme, the supply voltage or clock frequency can be adaptively changed at run-time; i.e, the circuit performance adapts itself based on the degradation. For example, Wang *et al.* proposed in [160] a novel technique to accurately evaluate the aging at run-time. They use monitors to examine the aging of the critical reliability paths accurately without impacting the normal operation of the chip.

Thereafter, the circuit is fine-tuned with either adaptive body biasing, and supply voltage or dynamic frequency scaling. Mintarno *et al.* investigated in [161,162] a scheme for dynamic control optimization of different self-adjusting parameters such as dynamic cooling, supply voltage, and clock frequency over lifetime for the circuit aging. Oboril *et al.* presented in [163] a dynamic run-time approach using voltage and frequency scaling which depend on run-time monitoring of temperature, performance, power and aging. Kumar *et al.* proposed in [164] to make use of the Adaptive Body Biasing and Supply Voltage techniques to realize a higher performance for an aged circuit.

- **Computational sprinting:** This mitigation scheme makes use of relaxation periods to recover from aging. (i.e., after fast execution of the task, it switches off for recovery); this approach is based on increased Vdd (higher than nominal value) for fast execution and power-gating for relaxation. For example, Gupta *et al.* presented in [165] a new technique for saving architectural power and reducing BTI induced degradations in digital circuits; they motivated this by the human circadian rhythm. Khoshavi *et al.* estimated in [166] the effectiveness of power-gating on aging and proposed a preferred *Sleep Transistor* configuration to achieve an improved aging mitigation approach. Chen *et al.* presented in [167] a comprehensive view of the existing optimization techniques and provided a design guideline on NBTI mitigation such as reducing the delay, tuning electrical parameters (voltage, V_{th}, and stress time), and presented leakage reduction approaches such as Input Vector Control, Internal Node Control, and Power Gating.

ADAPTIVE RESOURCE MANAGEMENT

In adaptive resource management, mitigation is used to balance the degradation between the available resources, e.g., via management of disposable resources. The adaptive resource management mitigation schemes can be divided into three types. They are Idle-Time Leveraging scheme, controlled resource wear-out, and spatial redundancy. The research in these areas is described next.

- **Idle-Time Leveraging (ITL) scheme:** The ITL is a mitigation scheme that advantageously use the idle time to mitigate the degraded device. For example, Calimera *et al.* proposed in [168] a new technique for aging-aware power gating by powering down standby states. Oboril *et al.* proposed in [169] ExtraTime: this is an aging-aware micro-architectural framework that can be used in the initial design stages for design space exploration. The proposed approach aims not only to enhance performance and power but also to mitigate aging without having detailed knowledge about the final hardware implementation. Oboril *et al.* presented in [170] a cross-layer technique that integrates the circuit, micro-architecture, and workload information to efficiently compensate device aging; and it uses new aging-aware instruction scheduling (i.e., classified as worst case and best case instructions). At the execution stage, the instructions use their specialized functional unit and this increases the idle ratio of the units executing the workloads. Hence, extending the lifetime when compared to balanced scheduling scheme. Firouzi *et al.* presented in [171] an efficient input vector selection methodology based on linear program-

ming to co-optimize the NBTI-induced delay degradation and to reduce leakage power consumption during standby mode.

- **Controlled resource wear-out:** This is a mitigation scheme which is based on dynamic supply voltage aging adjustment technique in other to control available many core resources on a chip; at the same time increasing the performance and reducing the power consumption. For example, Karpuzcu *et al.* presented in [172] the Dynamic Voltage Scaling for Aging Management (*DVSAM*) - a technique for managing processor aging (i.e., slowing down aging by reducing the operating parameters such as supply voltage and temperature) to operate in high performance or low power mode. Furthermore, it proposed the Bubble-Wrap many-core, a new architecture that extensively uses the *DVSAM* to reduce the many-core power wall. The *DVSAM* selects power efficient cores which are known as *throughput* cores while power inefficient cores are known as *expendable* cores. The *expendable* cores are worn out by subjecting them into high stress while preserving *throughput* core. As a result the Bubble-Wrap reaches highest sequential acceleration.

- **Spatial redundancy:** This is a mitigation scheme where components are duplicated. In case a failure occurs, the redundant unit is used instead; thereby compensating or increasing the lifetime of the device. For example, Srinivasan *et al.* studied in [173] two approaches (i.e., Structural Duplication i.e., continuous functionality while redundant structures are failed and switched off) that leverage micro-architectural structural redundancy for lifetime reliability enhancement. Ashraf *et al.* proposed in [174] a circuit-level technique known as Logic-Wear-Leveling which uses Dark-Silicon to mitigate the impact of BTI in logic data-paths. It introduced fine-grained spatial redundancy in timing vulnerable logic components, and leverages it at run-time to enable post-Silicon adaptability.

In addition to the aging adaptation and mitigation techniques explained above, it is worth to note that there are other mitigation methodologies for digital systems at higher abstraction level for functional errors. Examples are hardware platform, and mapping and software platform techniques [175].
All the presented mitigation techniques focus mostly on logic circuits or systems (*i.e., processor or architecture*). However, most of these mitigation techniques can be applied to the memory systems as well; and in particular the memory peripheral circuitry which is the main focus of this dissertation.

1.5. CHALLENGES

As already mentioned, reliability failure mechanisms give rise to serious challenges that may even prevent the deployment of nanotechnologies for critical applications and / or long lifetime application. Challenges should be addressed and appropriate solutions should be found. In case of embedded memories, which is the topic of this thesis, the reliability challenges can be classified into two classes:

- **Memory reliability modeling and characterization:** These are very important in determining the extent in which the memory is affected by aging, and how it de-

pends on workloads, technology, design, process, voltage and temperature variations. Therefore, memory reliability modeling and characterization are faced with a couple of challenges such as the following:

  - One of the principal question is *how to accurately model and characterize aging failure mechanisms.* Accurate modeling and characterization of the failure mechanism is crucial in predicting the reliability of the memory system.
  - The memory systems is made up of different components such as cell array, address decoders, sense amplifiers, write drivers, control circuit, pre-charge circuit and so on. Currently, research mostly focuses on the cell array. Therefore *analyzing the impact of aging with respect to all individual components and the complete memory including the interaction between its components still needs to be analyzed.* This analysis has to be done while considering process variation, temperature, supply voltage, and workloads.
  - Technology scaling typically worsens the aging impact. Therefore, an important task is *to analyze the severity of impact of aging while considering newer technology nodes as well as new device technologies such as FinFET.*
  - Finally, in the literature, aging was mostly analyzed for the 6T cell design. However, there are different cell designs such as 4T, 8T, etc. *Analyzing the impact of aging for different cell designs could lead to different results.*

- **Memory mitigation schemes:** Designing for reliability using mitigation schemes is an interesting research topic both in the industry and the academia. The reliability of an electronic system, in particular a memory system is crucial not only to the designer but also to the end user (consumer). As the modeling approaches were not accurate, existing mitigation schemes may not be appropriate. The main problem is that the modeling approaches were not incorporating all relevant effects, and that especially the workload dependence was not incorporated. So instead of first looking at the mitigation, the modeling had to be revisited first! Once this is done, appropriate mitigation schemes can be developed by answering the following:

  - The first key question is *how to use the information obtained from* ***accurate memory reliability modeling and characterization*** *to appropriately determine which mitigation scheme has to be selected. This can be a static, dynamic or hybrid scheme.*
  - The second key question is *how to evaluate the impact of the selected mitigation scheme.* Proper evaluation is required to ensure the reliability of the system.

1.6. RESEARCH TOPICS

The conducted research in this dissertation can be divided into three main parts. It is worth noting that most of the work done in this dissertation centered on the impact of BTI on memory sense amplifier designs, while few work is on impact of aging and mitigation of memory read-path, and write-path, respectively.

1. Analysis of aging impact on different memory sense amplifiers.
2. Investigation of aging impact and mitigation on memory read-path.
3. Investigation of write-path aging.

ANALYSIS OF AGING IMPACT ON DIFFERENT MEMORY SENSE AMPLIFIERS

In the past, several works concentrated on the impact of BTI on the cell array while not much work has been done with respect to the impact of BTI on memory sense amplifier. Moreover, SRAM sense amplifier is a critical component of a memory system which ensures that correct data is read out from the cell within a reasonable time due to its amplification nature.

Therefore, investigating and understanding the impact of BTI models on sense amplifier should be explored. In this dissertation, we focus on analyzing the impact of BTI on memory sense amplifier while considering various SA designs, technology nodes, process variations, supply voltages, temperatures, and workloads (*applications*).

INVESTIGATION OF AGING IMPACT AND MITIGATION ON MEMORY READ-PATH

There are different concerns by researchers with respect to the impact of BTI to the memory systems. Some argue that either memory cell array or peripherals circuits (such as SA, write driver etc.) are the most susceptible to aging impact. However, memory system consists of two major operations (i.e., read and write). These two operations create two important path to the memory system which are write-path and read-path. It is known that memory is more sensitive to reads than writes. In this dissertation, we focus on analyzing the aging impact of memory read-path while considering all memory components in that path, the way they interact with each other with respect to different supply voltages, temperatures, workloads, and technology nodes.

Another important research question is how to mitigate the most susceptible memory components of the read-path to aging. Therefore, it is vital to develop mitigation techniques to compensate for the aging effect on memory read-path. In this dissertation, we focus on developing two mitigation schemes for the memory read-path while considering its individual components and when combined together for various workloads.

INVESTIGATION OF WRITE PATH AGING

Memory write driver is responsible to ensure that correct data is written to the memory cell. Therefore, it is worth to investigate how aging does impact the memory write path (write driver). In this dissertation we focus on effectively characterizing and analyzing the actual impact of aging on write driver while taking into account various supply voltages, temperatures, and technology nodes.

1.7. CONTRIBUTIONS

THE contributions of this dissertation is directly associated to the research topics presented in the preceding section:

1

1.7.1. ANALYSIS OF AGING IMPACT ON DIFFERENT MEMORY SENSE AMPLIFIERS

Several investigations are carried out on the impact of BTI on memory sense amplifiers with respect to different BTI models, SA designs, technology nodes, and workloads (applications). With regard to this work, the contributions are made.

1. We investigated thoroughly and quantitatively the impact of BTI by utilizing various workloads and supply voltages for 45 nm technology node for both Reaction-Diffusion (R-D) and Atomistic trap-based models on SRAM standard latch sense amplifier [176].

2. We investigated the impact of BTI on the drain-input latch sense amplifier's sensing delay and sensing voltage while considering different technology nodes such as 90, 65, and 45 nm. In addition, we investigated the compensation techniques for the BTI degradation on sense amplifier using supply voltage [177].

3. We analyzed the impact of BTI on the sense amplifier's sensing delay by analyzing the BTI stress and relaxation cycles for each individual transistor to obtain more accurate results. These stress and relaxation cycles are workload dependent and we defined eight realistic workloads. In addition, we investigated the impact of BTI on the SA for deeply nano-scaled technology nodes such as 45, 32, 22, and 16 nm under different supply voltages, temperatures and workloads [178].

4. We investigated the impact of BTI on the standard latch-type SA design as a result of its superior performance. In addition, we analyzed the impact on process, voltage, and temperature (PVT) variations in combination with BTI for different workloads and technology nodes [179].

5. We investigated an accurate technique to quantify the impact of variability on the SRAM sense amplifier offset-voltage, while taking into account both process and time-dependent variations. To our knowledge, we are the first to determine the SA offset in the presence of all kinds of variability. The used method is accurate in the sense that it uses the Atomic Model for aging (which is a calibrated model) and considers the workload (as the aging variations are strongly workload dependent). Guaranteeing a resilient SA needs not only a correct sensing delay, but also an appropriate offset-voltage during the memory operational lifetime [180]. In addition, we investigated the sensitivity analysis of the SA at time zero. Then, we investigated the SA offset voltage specification while considering five global process corners at time-zero. Moreover, we analyzed the SA offset voltage specification for both time-zero and time-dependent variations on various supply voltages, temperatures and workloads while taking into account various process corners. Finally, we analyzed the failure rate for the offset voltage specification at nominal process corner while considering various workloads, supply voltages and temperatures.

6. We performed a comparative study of aging analysis while taking into account different supply voltages and temperatures for three memory Sense Amplifier designs

such as Low Power (LP), Mid Power/Performance (MP), and High Performance (HP). The evaluation metric, the Sensing Delay (SD) of the three designs is analyzed for various workloads using 45 nm technology. Furthermore, the SA tuning technique is adopted to determine the SA offset specification of all SAs. In addition, we performed this analysis using the realistic atomistic model calibrated with representative test data [181].

1.7.2. INVESTIGATION OF AGING IMPACT AND MITIGATION ON MEMORY READ-PATH

With regard to the aging analysis and mitigation of the SRAM read path. We investigated the impact of BTI on the read path of 32 nm high performance SRAM. It integrates the impact on the memory cell, the sense amplifier, and the way they communicate to each other. The investigation is done while taking into account various workloads and by examining both the bit-line swing which reveals the degradation of the cell and the sensing delay which reveals the degradation of the sense-amplifier. The voltage swing on the bit lines has a direct impact on the proper performance of the sense amplifier [182].

In addition, we proposed an appropriate technique to predict and mitigate the effect of BTI on the read path of a high performance SRAM design. We also investigated the effect of the memory cell, and Sense Amplifier (SA), and the way they communicate with each other. This technique evaluates various workloads, and technology nodes while examining both the bit-line swing and the sensing delay.

1.7.3. INVESTIGATION OF WRITE PATH AGING

With regard to memory write driver, we investigated of the impact of aging on the write driver's average write delay and its distribution [183]. The analysis is performed at nominal supply voltage for different technology nodes of 45, 32, and 22 nm, respectively; while taking into account both nominal and high temperature such as 298K and 398K, respectively. The sensitivity to supply voltage fluctuations and temperatures are also explored.

1.8. THESIS ORGANIZATION

THE remainder of this dissertation is organized as follows.

Chapter 2 provides the contributions of this dissertation with respect to the analysis of aging impact on various memory sense amplifiers. First, it presents the comparative study of both RD and Atomistic Trap-Based BTI models used SRAM Sense Amplifier reliability analysis. Second, it gives the BTI impact on the drain-input latch type sense amplifier for different supply voltages and technology nodes. Third, it presents combined impact of BTI and supply voltage, temperature variation on standard latch type sense amplifier. The latter work is extended to cover the impact on process, voltage, temperature together with aging, different workloads and technology nodes. Fourth, it presents the proposed accurate technique to quantify the impact of variability on offset voltage of the standard latch type sense amplifier while considering both time-zero and time-

dependent variations. In addition, the latter is extended to analyze the sensitivity of the sense amplifier at time-zero, failure rate analysis and their combined impact (time-zero and time-dependent) while different process corners. Finally, it presents the comparative study of the impact of aging on different sense amplifier designs while taking into account various supply voltages and temperatures.

Chapter 3 provides the contributions of this dissertation with respect to the read path aging. It presents the impact of BTI on a high performance SRAM, and the BTI impact on the individual parts of memory design (i.e., the memory cell and the SA design), and how they interact with each other. The evaluation considered metrics for the memory cell and the sense amplifier are the sensing delay (SD) and the Bit-line swing (BLS).

Chapter 4 provides the contributions of this dissertation with respect to the write path aging. It presents the BTI impact on the memory write driver, while taking into account various technology nodes, supply voltages and temperatures. The evaluation metric used for the write driver is the write delay.

Finally, the conclusion and future work are given in Chapter 5.

2

BTI IMPACT ON DIFFERENT MEMORY SENSE AMPLIFIERS

The content of this chapter incorporates the following research articles:

1. ***I. O. Agbo**, M. Taouil, S. Hamdioui, S. Cosemans, P. Weckx, P. Raghavan, F. Catthoor, Comparative analysis of RD and Atomistic trap-based BTI models on SRAM Sense Amplifier, 2015 10th International Conference on Design Technology of Integrated Systems in Nanoscale Era (DTIS), pp. 1–6, April 2015, Naples, Italy.* ***Best Paper Award***

2. ***I. O. Agbo**, S. Khan, S. Hamdioui, BTI impact on SRAM sense amplifier, 2013 8th IEEE Design and Test Symposium, pp. 1–6, Dec. 2013, Marrakesh, Morocco.*

3. ***I. O. Agbo**, M. Taouil, S. Hamdioui, H. Kükner, P. Weckx, P. Raghavan, F. Catthoor, Integral impact of BTI and voltage temperature variation on SRAM sense amplifier, IEEE 33rd VLSI Test Symposium (VTS), pp. 1–6, April 2015, Napa, USA.*

4. ***I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, H. Kükner, P. Weckx, P. Raghavan, F. Catthoor, Integral Impact of BTI, PVT Variation, and Workload on SRAM Sense Amplifier, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 1444–1454, volume 25, issue 4, April 2017.*

5. ***I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, Quantification of Sense Amplifier Offset Voltage Degradation due to Zero-and Run-Time Variability, 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 725–730, July 2016, Pittsburgh, USA.* ***Best Paper Award***

6. ***I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, Sense Amplifier Offset Voltage Analysis for both Time-zero*

and Time-dependent Variability, ACM Transactions on Design Automation on Electronic Systems (TODAES), under review.

7. ***I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, Comparative BTI analysis for various sense amplifier designs, 2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS), pp. 1–6, April 2016, Košice, Slovakia.*

2015 10th International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS)

Comparative Analysis of RD and Atomistic Trap-Based BTI models on SRAM Sense Amplifier

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Stefan Cosemans Pieter Weckx Praveen Raghavan Francky Catthoor
IMEC vzw
Kapeldreef 75, 3001 Leuven, Belgium
{stefan.cosemans, Pieter.Weckx, Ragha, Francky.catthoor}@imec.be

***Abstract*—Bias Temperature Instability (BTI) in transistors has become a key reliability bottleneck with sub-45nm CMOS technologies. The most common models to characterize BTI are the Reaction-Diffusion (RD) and Atomistic trap-based models. This paper presents comparative impact analysis of RD and Atomistic trap-based BTI models for the SRAM Sense Amplifier. The evaluation metric, the sensing delay is analyzed for both models for the different workloads and supply voltages for 45nm technology node. The results show that the sensing delay degradation is slightly higher in RD model than Atomistic trap-based model for different workloads. Nevertheless, we observe a similar trend for both models. For example the BTI impact degradation is 6.69% for RD model and 6.57% for Atomistic trap-based model when worst case workload is applied for a 10^8s life time.**

***Index Terms*—BTI, NBTI, PBTI, SRAM sense amplifier**

I. INTRODUCTION

In recent decades, CMOS technology has been sustained with aggressive downscaling that severely impacts the reliability of devices [1,2,26]. These trends are a consequence of advancements in the fabrication technology, introduction of novel materials and evolution of architecture designs. Bias Temperature Instability (BTI) (i.e., Negative BTI in PMOS transistors and Positive BTI in NMOS transistors) is a reliability failure mechanism which affects the performance of MOS transistors by increasing their threshold voltage and reducing their drain current (I_d) over the operational lifetime [4,21]. However, studies of such individual devices or small composites do not allow extrapolation of these effects on larger circuits like SRAMs.

Static Random-Access Memories (SRAM) occupy a large fraction of semiconductor chip and play a major role in the silicon area, performance, and critical robustness [13]. An SRAM system consists of an array of cells, its peripherals circuits such as row and column address decoders, control circuits, write drivers, and sense amplifiers.

In recent years, there has been an increased attention in modeling BTI failure mechanisms. There are two well known BTI models (i.e., Reaction-Diffusion (RD) models and Atomistic trap-based models) in the field of aging failure mechanisms. Several work has been published on RD and Atomistic models. For instance, Zafar [21] presented BTI as a statistical mechanism based model for negative bias temperature instability induced degradation. Alam [22] presented a critical examination of the mechanics of dynamic NBTI for PMOS. Kaczer et al. [18] proposed Atomistic trap-based model which model BTI and Random Telegraph Noise (RTN) behavior for Sub-45nm devices with workload dependencies. Both models have been independently used at the gate level by researchers. Khan et al. [16] presented BTI analysis of different gates for the dynamic inputs using RD models while including the periodic waveforms only. Kukner et al. [23] analyzed BTI impact on a single inverter using Atomistic trap-based model. However, very limited work is done in comparing both BTI models. Kukner et al. in [19] presents comparison of RD and Atomistic trap-based BTI models for logic gates only while no comparison of the two models has been reported in literature at circuit levels. Further also, Agbo et al. in [5] analyzed an integral impact of BTI and voltage temperature variation on SRAM sense amplifier using RD model while including different workloads, supply voltages and temperatures. Again, Agbo et al. in [20] investigated BTI analysis for high performance and low power SRAM sense amplifier designs using Atomistic model while considering different SA designs. However, comparative analysis of RD and Atomistic trap-based model on circuit level (including sense amplifier) while considering different workloads and supply voltages still needs to be explored. It is worth noting that comparative study of both models at circuit level will help in understanding and selection of the best suited model for quantifying the aging rate of each memory parts required for optimal reliable memory design. This paper focuses on the comparative study of two different BTI models (i.e., RD and Atomistic trap-based models) on SRAM sense amplifiers each targeted for a different application. The standard latch-type sense amplifier design is selected for its superior performance and industrial representativeness [17]. The comparison of the two BTI models impact for standard latch sense amplifier is analyzed using different workloads and supply voltages. The main contributions of the paper are as follows.

- Investigation of BTI impact on the sense amplifier's sensing delay, two different target BTI models are considered.
- Thorough quantitative analysis of the BTI impact using different workloads for 45nm technology node for the two BTI models are investigated.
- Analysis of BTI impact under different supply voltages for SRAM sense amplifier sensing delay using different

978-1-4799-1999-4/15/$31.00 ©2015 IEEE

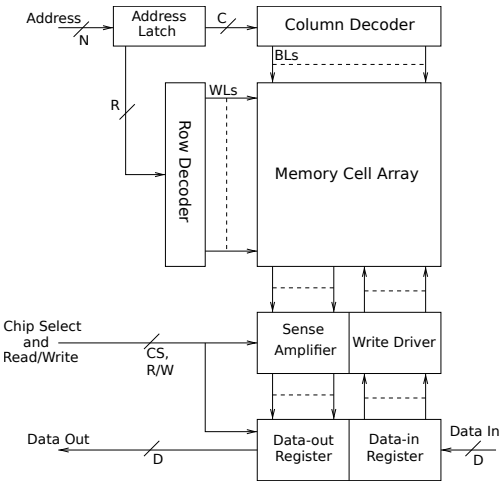


Fig. 1. Functional model of SRAM system

workloads for both Reaction-Diffusion (R-D) and Atomistic trap-based models are explored.

- Comparison between RD and Atomistic model for the above mentioned.

The rest of the paper is organized as follows: Section II introduces the sense amplifiers and both BTI models. Section III provides our analysis framework, it presents also the performed experiments. Section IV analyzes the results for comparing both models for different workloads and supply voltages. Finally, Section V concludes the paper.

II. BACKGROUND

This section presents the working principles of the targeted sense amplifier. Thereafter, it explains the Reaction Diffusion and Atomistic trap-based BTI models analyzed in this paper.

A. Memory model

Figure 1 shows a functional model of the SRAM system [26]. A memory system is comprised of a memory cell array, row and column address decoders, read/write circuitry, input/output data registers and control logic. The main target of this paper is the comparative investigation of Atomistic and RD model on sense amplifier.

B. SRAM Sense Amplifier

Several implementations of sense amplifiers have been proposed. In this section, first the standard latch-type SRAM strobed sense amplifier will be addressed which is representative for industrial SA designs [17].

Standard Latch-Type Sense Amplifier (SLT SA)

A sense amplifier (SA) in SRAMs is responsible for the amplification of a small voltage difference at the bit lines (i.e., BL (BLBar)) during read operations.

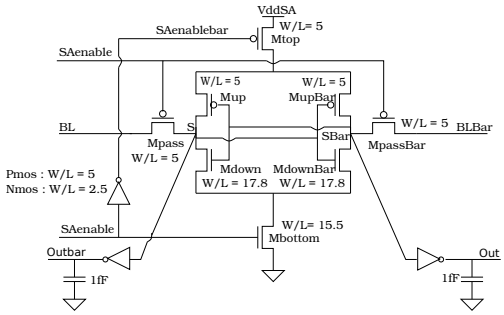


Fig. 2. Standard latch-type Sense Amplifier

The structure of the Standard latch-type Sense Amplifier is depicted in Fig. 2. The width length ratio of each transistor is presented by W/L.

The operation of the sense amplifier comprises of two phases. In the first phase, when SAenable is low, the access transistors Mpass and MpassBar connect to the bitlines (i.e., BL and BLBar) with the internal nodes S (SBar). In this phase, Mtop and Mbottom transistors are switched off. In the second phase, when SAenable is high, the pass transistors disconnect the BL (BLBar) input from the internal nodes. The cross coupled inverters get their current from Mtop and Mbottom and subsequently amplify the difference between S and SBar and produce digital outputs on Out and Outbar. S (SBar) node is actively pulled down when SBar (S) exceeds the threshold voltage of Mdown. The positive feedback loop ensures low amplification time and produces the read value at its output. Moreover, all current paths are disabled when S (SBar) is at 0V and SBar (S) is at $V_{dd\mathrm{SA}}$ or vice versa. This process is repeated for each read operation.

C. Reaction Diffusion Model

The Bias Temperature Instability (BTI) mechanism takes place inside the MOS transistors and causes a threshold voltage shift that impacts the delay negatively; its mechanism is described below.

BTI Mechanism

BTI increases the absolute $\mathrm{V_{th}}$ value in MOS transistors. For the PMOS, the negative $\mathrm{V_{th}}$ is further lower while for NMOS the $\mathrm{V_{th}}$ increases. The increment in the absolute value of the $\mathrm{V_{th}}$ in a PMOS transistor that occurs under *negative* gate stress is referred to as NBTI, and the one that occurs in an NMOS transistor under *positive* gate stress is known as PBTI. For a MOS transistor, there are two BTI phases, i.e., the stress phase and the relaxation phase.

Recently, exhaustive efforts have been put to understand NBTI [4,7,21]. Kaczer *et al.* in [7] have analyzed NBTI using an atomistic model. Alam *et al.* [4] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion process. The model is usable at a higher level such as circuit

level. In Kukner *et al.* in [19], both the atomistic and RD models have been compared. The authors conclude that for the long-term simulation time, RD model is lightweight than the atomistic model. For this reason we select this model [4].
Stress Phase: In the stress phase, the Silicon Hydrogen bonds (≡Si-H) break at Silicon-Oxide interface. The broken Silicon bonds (≡Si-) remain at the interface (known as interface traps), and the released H atoms/molecules diffuse towards the gate oxide. The number of interface traps (N_{IT}) generated after applying a stress of time (t) is given by [4]:

$$N_{IT}(t) = \left(\frac{N_o \cdot k_f}{k_r}\right)^{2/3} \cdot \left(\frac{k_H}{k_{H_2}}\right)^{1/3} \cdot (6 \cdot D_0 \cdot t)^{1/6} \quad (1)$$

where N_o, k_f, k_r, k_H, and k_{H_2}, represent initial ≡Si-H density, ≡Si-H breaking rate, ≡Si- recovery rate, H to H_2 conversion rate, and H_2 to H conversion rate inside the oxide layer, respectively. D_0=D_{H_2}.exp($-E_A$/kT) [11] is the diffusion coefficient of the produced H_2 species and E_A is the activation energy, k is the boltzman constant, and T is the temperature in Kelvin.

Relaxation Phase: In the relaxation phase, there is no ≡Si-H breaking. However, the H atoms/molecules diffuse back towards the interface and anneal the ≡Si- bonds. The number of interface traps that *do not* anneal by the approaching H atoms during the relaxation phase is given by [15]:

$$N_{IT}(t_o + t_r) = \frac{N_{IT}(t_o)}{1 + \sqrt{\frac{\xi \cdot t_r}{t_o + t_r}}} \quad (2)$$

where $N_{IT}(t_o)$ is the number of interface traps at the start of the relaxation, ξ is a relaxation coefficient with ξ=0.5 [15], t_o is the duration of the previous stress phase and t_r is the relaxation duration.

Threshold voltage increment: The N_{IT} oppose the gate voltage which result in a threshold voltage increment (ΔV_{th}). The relation between N_{IT} and ΔV_{th} is given by [8]:

$$\Delta V_{th} = (1 + m) \cdot q \cdot N_{IT}/C_{ox} \cdot \chi, \quad (3)$$

where m, q, and C_{ox} are the holes/mobility degradation that contribute to the V_{th} increment [9], electron charge, and oxide capacitance, respectively. χ is a BTI coefficient with a value χ=1 for NBTI and χ=0.5 for PBTI [10].

D. Atomistic Model

Kaczer et al. proposed the atomistic model in [18,19]. It is based on the capture and emission of single traps during stress and relaxation phases of NBTI/PBTI respectively. The threshold voltage shift of the device ΔV_{th} is the accumulated results of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by [24]

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c + \tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e} + \frac{1}{\tau_c})t_{STRESS}\right]\right\} \quad (4)$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c + \tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e} + \frac{1}{\tau_c})t_{RELAX}\right]\right\} \quad (5)$$

where τ_c and τ_e are the mean capture and emission time constants, and t_{STRESS} and t_{RELAX} are the stress and relaxation periods, respectively. [25], gives insight for the relation between (1), (2), and the V_{th}.

III. ANALYSIS FRAMEWORK

A. Framework Flow

In order to evaluate the BTI impact of both models, we replicate the flows of [5] for RD and [6] for Atomistic model. Next, the generic inputs of both models are described.
Generic input: The general input blocks of the framework are the technology library, Sense Amplifier design, and BTI input parameters.

- Technology library: In this work we only use the 45nm PTM library [27]. Note that in general any library card can be used.
- SA design: Generally, all sense amplifier design can be used. In this paper we focus only on the standard latch-type SA. The SA design is described by a SPICE netlist.
- BTI parameters: The BTI induced degradation depends strongly on the stress time duration. The stress time defines how long the workload sequence is being applied. The workload sequence is assumed to be repeated until the age time is reached. To perform realistic workload analysis, we assume that today's application consists of 10% - 90% memory instructions and the percentage of read instructions is typically 50% - 90%. Furthermore, we derive from this the following cases: best case with a stress period of 0.1 * 0.1 = 0.01, worst case with 0.9 * 0.9 = 0.81, and mid case with 0.5 * 0.5 = 0.25. They lead to the following workload sequences: best-case: $R0R1I^{198}$, worst-case: $R0^4I^1$, mid-case: $(R0)I^{24}$. In these sequences, $R0$ stands for read 0, $R1$ stands for read 1, I for idle operation (which includes memory write operations).

Processing: Here, we described shortly the flows of [5] RD model and [6] Atomistic model.

- RD: There are two processing blocks, the BTI predictor and the HSPICE simulation unit. The long term BTI predictor uses the duty factor, frequency, and aging to predict the interface traps/ threshold voltage increment of each device in the sense amplifier. In addition to workload inputs, inputs are required from the reaction-diffusion model (such as K_f, K_r, D_H, etc.), technology parameters, and voltage temperature (VT). Once the BTI induced V_{TH} increments are calculated per transistor, the original BTI free netlist will be updated. This new netlist is simulated in HSPICE/Verilog-A. **Output:** Finally, post-analysis of the results are performed for varying voltages and temperatures in MATLAB environment.

- Atomistic: Based on the transistor dimensions and the other specified inputs, a Control script (perl) generates several instances of BTI augmented SRAM sense amplifier circuits. Every generated instance has a distinct

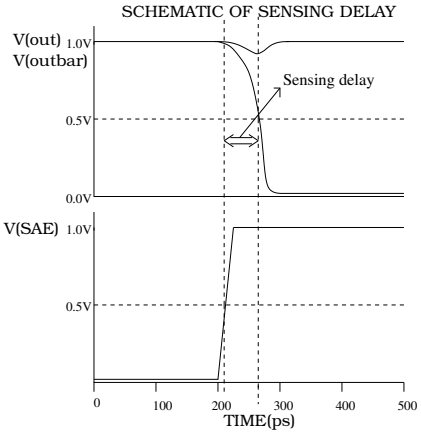


Fig. 3. Metric diagram of Sensing delay.

number of traps (with their unique timing constants) in each transistor, and are incorporated in a Verilog-A module of the SA netlist. The module responds to the every individual trap, and alters the transistors concerned parameters such as V_{th}. After inserting BTI in every transistor of the SA design, a Monte Carlo (MC) is performed at different time steps (100 runs at each time step) where circuit simulator (HSPICE/Spectre) is used to investigate the BTI impact.

B. Output Metrics

In this section, the sensing delay metric used for analyzing BTI impact on sense amplifier is described.

Sensing delay: The sensing delay metric is determined when the trigger signal (i.e.; sense amplifier enable input signal) reaches 50% of the supply voltage and the target (i.e.; either out or outbar falling output signal) reaches 50% of the supply voltage. The sensing delay is defined as the time difference between the target and the trigger as shown in Fig. 3.

C. Experiments Performed

In this paper, three sets of experiments are performed to analyze BTI impacts. These experiments are described below:

1 **Temporal Impact Experiments:** BTI impact on sensing delay for the two models on Sense Amplifier design is investigated for different periods of stress.
2 **Workload Dependent Experiments:** BTI impact on the sensing delay for the two models using three workloads is investigated.
3 **Supply Voltage Dependent Experiments:** BTI impact on the sensing delay of the SRAM sense amplifier for varying supply voltages (i.e., from -10% of V_{dd} to V_{dd} and +10% of V_{dd}) for two workloads (i.e., worst case and best case) and two BTI models (i.e., Atomistic trap-based and Reaction Diffusion) is explored.

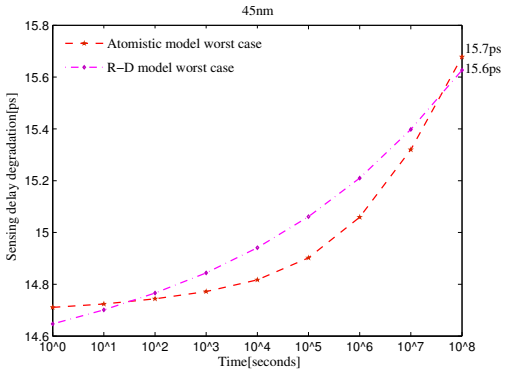


Fig. 4. Reaction Diffusion and Atomistic model BTI impact on sensing delay.

IV. EXPERIMENTAL RESULTS

In this section, we present the analysis results of the experiments mentioned in the previous section. BTI affects the sensing delay of the sense amplifier, i.e., the time required to amplify the input from BL and BLBar to outputs *out* and *outbar* (see Figures 1 and 2). In order to quantify this delay, we simulate the initial BTI-free SA design, for 45nm technology node and take its sensing delay as reference. To obtain proper sensing delays, appropriate values of BL and BLBar should be selected. For 45nm, we assume the differential input to be 89.4mV (V_{dd}) [17].

A. Temporal Impact Experiments

Figure 4 shows the relative increment of the sensing delay w.r.t. the stress time (aging) for worst-case workload for both Atomistic model and R-D model. Both models show a similar trends. For example, for an operation time between 10^0sec and 10^8sec, the delay increases from 14.7ps to 15.7ps for Atomistic model approximately 6.6% and approaches 14.7ps to 15.6ps for R-D model which is 6.7%.

B. Workload Dependent Experiments

The BTI induced degradation is sensitive to the workload. The workload defines when and how long each transistor is stressed. Figure 5 shows the BTI impact on sensing delay for both Atomistic and RD models. For the Atomistic model, the BTI sensing delay degradation for worst case, mid case, and best case equals 14.71ps, 14.67ps and 14.64ps at 10^0s stress, respectively. At 10^8s they equal to 15.69ps, 15.13ps and 14.85ps, respectively. For RD model, these values are 14.65ps, 14.55ps, and 14.50ps at 10^0s and 15.63ps, 15.08ps, and 14.76ps, at 10^8s, respectively. Both models observe the same trends. However, the relative numbers may differ. For example, the relative sensing delay increment equals 6.57% from 10^0s to 10^8s for Atomistic model when worst case is applied. However, for RD model, this degradation at 10^0s worst case is equal to 14.65ps and approaches 14.55ps and 14.50ps for mid case and best case, respectively.

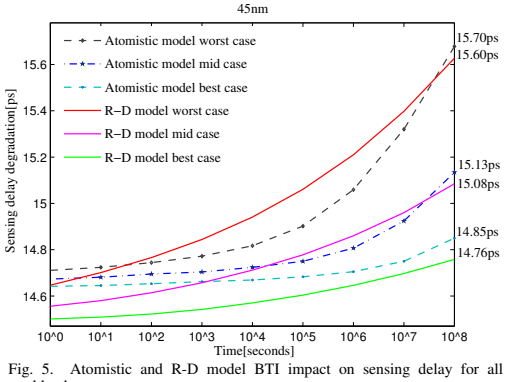


Fig. 5. Atomistic and R-D model BTI impact on sensing delay for all workloads.

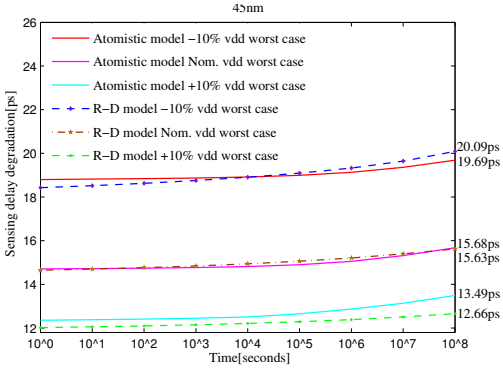


Fig. 6. Atomistic and R-D model BTI impact on varying supply voltage for worst case workload.

There is also significant BTI impact variation for Reaction Diffusion model. For instance, the BTI impact variation is equal to 6.69% for worst case, and approaches 3.63% and 1.78% for mid case and best case, respectively. When comparing the two model w.r.t. sensing delay, in the degradation free case, Atomistic model is 14.71ps for worst case while R-D model is 14.65ps for worst case workload which is 0.41% and approaches 0.76% and 0.97% for the mid case and best case, respectively. There is significant change relatively in the BTI induced degradation for the two models. For instance, Reaction Diffusion model is worse than Atomistic model with 0.12% for the worst case and approaches 0.49% and 0.36% for the mid case and worst case, respectively. Moreover, the two models maintain the same trends for the three workloads considered.

C. Supply Voltage Dependent Experiments

Supply voltage fluctuations generally impact the operating condition of MOS transistors. Supply voltage variations in a transistor can impact the sensing delay significantly as it impacts the operational speed. In addition, variation in supply

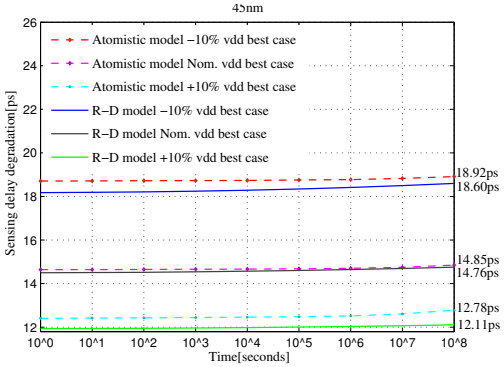


Fig. 7. Atomistic and R-D model BTI impact on varying supply voltage for best case workload.

voltage also affects the oxide field (capacitance) and subsequently the BTI impact (see C_{ox} in Eqn. 3) for RD model. The analysis of the supply voltage variation is performed for both BTI models (i.e., Atomistic and Reaction Diffusion models) and two workloads, (i.e., worst case and best case) for 45nm technology. The supply voltage is varied between -10% and +10% of nominal V_{dd}, i.e., between 0.9V and 1.1V. Figures 6 and 7 depict the BTI induced sensing delay for various supply voltages and BTI models for the worst case and best case workloads, respectively. From the figures we conclude the following:

- Increasing the supply voltage reduces the sensing delay degradation. For instance, in Figure 6, for an Atomistic trap-based model, the BTI degradation at 10^0s is 12.36ps, 14.71ps and 18.80ps for $+10\%$ V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. At 10^8s, the sensing delay equals 13.49ps, 15.68ps and 19.69ps for $+10\%$ V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. The absolute numbers of varying supply voltage for Atomistic trap-based model maintain the claim of increase in supply voltage reduces the sensing delay degradation. The relative increment of the varying supply voltage for Atomistic trap-based model shows an opposite trend, for $+10\%$ V_{dd} the sensing delay increment equals 9.19% while 6.57% and 4.72% for nominal V_{dd} and -10% V_{dd}, respectively.

  The figure also shows, for Reaction Diffusion model, the BTI degradation at 10^0s is 12.02ps, 14.65ps and 18.43ps for $+10\%$ V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. Then, for BTI induced degradation at 10^8s is 12.66ps, 15.63ps and 20.09ps for $+10\%$ V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. The Reaction Diffusion model varying supply voltage absolute numbers also maintain the claim that increasing the supply voltage mitigates the sensing delay degradation. Furthermore, the relative numbers for R-D model varying supply voltages shows a consistent trend with the absolute numbers, for $+10\%$ V_{dd} the sensing delay is 5.26% while 6.69% and

9.03% for nominal V_{dd} and -10% V_{dd}, respectively.

- Figure 7 shows also increasing the supply voltage reduces the sensing delay degradation for best case workload. For instance, for an Atomistic model, e.g., +10% V_{dd}, the BTI degradation free case is 12.41ps while 14.64ps and 18.71ps for nominal V_{dd} and -10% V_{dd}, respectively. Furthermore, in the BTI induced degradation at 10^8s, +10% V_{dd} = 12.78ps and approaches 14.85ps and 18.92ps for nominal V_{dd} and -10% V_{dd}, respectively. The absolute numbers of the best case maintains the trend that increasing the V_{dd} mitigates the sensing delay degradation. However, this trends relatively is reversed, for e.g., +10% V_{dd} is equal to 3.01% while 1.43% and 1.10% for nominal V_{dd} and -10% V_{dd}, respectively.

  Besides, the figure shows varying supply voltages (i.e., -10% V_{dd}, Nom. V_{dd}, and +10% V_{dd}) for RD model best case. For instance, in BTI degradation at 10^0s is 11.94ps while 14.50ps and 18.18ps for +10% V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. Then, for BTI induced degradation at 10^8s is 12.11ps, 14.76ps and 18.60ps for +10% V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. For RD model best case varying V_{dd} absolute numbers sensing delay BTI impact keep to the trend that increasing the supply voltage reduces the sensing degradation. Moreover, the same trends is observed in the relative increment, for $+10\%$ V_{dd} is 1.45% while 1.78% and 2.34% for $+10\%$ V_{dd}, nominal V_{dd} and -10% V_{dd}, respectively. In conclusion, both models absolute numbers for worst case and best case workload varying supply voltage sensing delay are consistent with the trends, increase in supply voltages results in reduced BTI induced sensing delay degradation. However, both models show opposite trend relatively for increasing supply voltages in the presence of BTI induced degradation.

D. Discussion

Reliable and robust SRAM SA designs are crucial for the overall design of memory systems, and also to the design community. The current analysis focused on comparative study of both Atomistic trap-based and Reaction-Diffusion model on standard latch type SA for different supply voltages and workloads. Both models predict a similar reliability impact (in terms of delay) for the considered SA. We observed slightly higher sensing delay degradation in RD model than Atomistic model for various workloads. The V_{th} for Atomistic model is based on stochastic analysis. So the same circuit will result in a range of sensing delay increments, while for the RD model a single sensing delay increment value is used in our analysis. Therefore, based on the calibrations and the stochastic nature of Atomistic model, we conclude that it is best suited for BTI as it also includes worst-case and best-case analysis.

V. CONCLUSION

This paper investigated the combined impact of Bias Temperature Instability (BTI), different workloads and supply voltages on the standard latch type (SLT). The sensing delay degradation is more impacted by workload that contain more and longer stress periods and reduces with higher supply voltages. These trends have been observed by both the Reaction Diffusion and Atomistic trap-based BTI models.

REFERENCES

[1] ITRS, "International Technology Roadmap for Semiconductor 2004" "www.itrs.net/common/2004 update/2004update.htm.".

[2] S. Borkar, et al "Micro architecture and Design Challenges for Giga scale Integration", *Pro. of Intl. Sympos. Micro architecture*, 2004.

[3] S. Hamdioui et al, "Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes", *DATE*, 2013.

[4] M. A. Alam and S. Mahapatra, "A Comprehensive Model of PMOS NBTI Degradation", *Microelectronics Reliability* , Vol:45, 2005.

[5] I. Agbo, et al., "Integral Impact of BTI and Voltage Temperature Variation on SRAM Sense Amplifier" *VLSI Test Symposium* , 2015.

[6] S. Khan, et al, "Bias Temperature Instability Analysis of FinFET based SRAM cells", *DATE*, 2014.

[7] B. Kaczer, et al., "Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification", *IPRS* , pp: 381-387, 2005.

[8] B.C. Paul et al, "Impact of NBTI on the Temporal Performance Degradation of Digital Circuits", *IEEE Electron Device Letter*, Vol. 26, No.8, Aug. 2005.

[9] A. T. Krishnan, et al, "NBTI impact on transistor and circuit: Models, mechanisms and scaling effects", *IEDM*, 2003.

[10] M. T. Luque, et al, "From Mean Values to Distribution of BTI Lifetime of Deeply scaled FETs through Atomistic Understanding of the Degradation" *Sym. on VLSI Technology*, pp: 152-153, 2011.

[11] D. Varghese, et al, "On the Dispersive versus Arrhenius Temperature Activation of NBTI Time Evolution in Plasma Nitrided Gate Oxides: Measurements, Theory, and Implications", *IEDM*, Dec. 2005, pp. 1-4.

[12] S. Zafar, et al, "A comparative study of NBTI and PBTI in SiO2/HfO2 stacks with FUSI, TiN gates", *Pro. of VLSI Technology symp.*, 2006.

[13] P. Pouyan, et al, "Process Variability-Aware Proactive Reconfiguration Technique for Mitigating Aging effects in Nano Scale SRAM lifetime", *IEEE 30th VLSI Test Symposium.*, 2012.

[14] B. Cheng, A. R. Brown, "Impact of NBTI/PBTI on SRAM Stability Degradation", *IEEE ELECTRON DEVICES LETTERS*, 2011.

[15] S. Kumar, C.H. Kim, S. Sapatnekar, "Impact of NBTI on SRAM Read Stability and Design for Reliability", *ISQED*, pp: 212-128, 2006.

[16] S. Khan, et al, "BTI impacts on logical gates in nano-scale CMOS technology", *in Proc. IEEE Int. DDECS*, Apr. 2012, pp. 348 - 353.

[17] S. Cosemans, "Variability-aware design of low power SRAM memories", *Ph.D Thesis Katholieke Universiteit Leuven*, 2009.

[18] B. Kaczer, et al., "Origin of NBTI variability in deeply scaled pFETs", *Reliability Physics Symposium*, 2010.

[19] H. Kukner et al., "Comparison of Reaction-Diffusion and Atomistic Trap-based Models for Logic Gates", *IEEE transactions on device and materials reliability*, 2013.

[20] I. Agbo et al., "BTI Analysis for High Performance and Low power SRAM Sense Amplifier Designs", *MEDIAN*, 2015.

[21] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation", *J. Appl. Phys.,* vol. 97, no. 10, pp. 103709-1 - 103709-9, May 2005.

[22] M. Alam, "A critical examination of the mechanics of dynamic NBTI for pMOSFETs", *in IEDM Tech. Dig.,* , Dec, 2003, pp. 14.4.1-14.4.4.

[23] H. Kukner et al, "Impact of duty factor, stress stimuli, and gate drive strength on gate delay degradation with an atomistic trap-based BTI model", *in Proc. Euromicro DSD Conf. DSD*, Sep. 2012, pp. 1 - 7.

[24] M. Toledano-Luque et al, "Response of a single trap to AC Negative Bias Temperature Stress", *IRPS*, 2011.

[25] P. Weckx, et al, "Defect-based Methodology for Workload-dependent Circuit Lifetime Projections-Application to SRAM", *IRPS*, April 2013.

[26] S. Hamdioui, "Testing Static Random Access Memories: Defects, Fault Models and Test Patterns", *Kluwer Academic Press Publishers, The Netherlands*, 2004.

[27] Predictive Technology Model "http://ptm.asu.edu/",

BTI Impact on SRAM Sense Amplifier

Innocent Agbo Seyab Khan Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.S.K.Seyab, S.Hamdioui}@tudelft.nl

***Abstract*—Bias Temperature Instability (BTI) -Negative BTI in PMOS and Positive BTI in NMOS transistors- has become a key reliability bottleneck in the nano-scaled era. This paper presents BTI impact on SRAM's sense amplifier of different technologies, a robust sense amplifier has a lower sensing delay and higher sensing voltage. The results show that as technology scales down (i.e., from 90nm to 65nm, and 45nm), BTI impact on sensing delay increases, while that on the sensing voltage decreases, causing the sense amplifier memory, hence to be less robust and reliable. In addition, the paper also investigate the use of supply voltage to reduce the BTI degradation. The result show that increasing the power supply can reduce the sense amplifier BTI degradadtion with 33% for sensing voltage and with 18% for sensing delay; leading to clear tradeoff engineering question between power and robustness.**

***Index Terms*—BTI, NBTI, PBTI, SRAM sense amplifier**

I. INTRODUCTION

In recent decades, CMOS technology has witnessed relentless downscaling. Forces behind the trend are advancements in the fabrication technology, introduction of novel materials and evolution of architecture designs. However, for circuits based on the current CMOS technology, reliability failures have become a major bottleneck [1]–[3]. Bias Temperature Instability (BTI) (i.e., Negative BTI in PMOS transistors and Positive BTI in NMOS transistors) is a reliability failure mechanism which affects the strength of MOS transistors by increasing their threshold voltages and reducing their drain current (I_d) over the operational lifetime [5], [7], [8].

Static Random-Access Memories (SRAM) occupies a large part of semiconductor systems and plays a major role in the silicon area, performance, and critical robustness [9]. Much has been published on SRAM Test (e.g. [10]–[12]) than SRAM reliability (e.g. BTI) as reliability challenges emerged with technology scaling. Moreover, an SRAM system consists of cells array, and its peripherals circuits such as column and row address decoders, control circuits, write drivers, and sense amplifiers. Much have been published on the BTI SRAM cell array and few on the SRAM peripheral circuitry. For example, Binjie et al [17] investigated NBTI impact on Static Noise Margin (SNM) and Write Noise Margin (WNM) degradation of 6T SRAM cell. Kumar et al [18] Analyzed the impact of NBTI on the read stability and SNM of SRAM cells. Bansal et al [19] presented insights on the stability of an SRAM cell under the worst-case conditions and analyzed the effect of NBTI and PBTI, individually and in combination. On the other hand, few authors have focused on reliability analysis of the address decoders. For instance, Hamdioui et al. in [20] presented analysis of spot defects in SRAM address decoders and in [21] identified decoder delay faults due to inter and intra-gate resistive defects. Khan et al [22] investigated the impact of partial opens and BTI in SRAM address decoder. Furthermore, Menchaca et al [23] analyzed the BTI impact on different sense amplifier designs implemented on 32nm technology node by using failure probability (i.e., flipping a wrong value) as a reliability metric. However, the impact of BTI for different technology nodes and varying supply voltages are yet to be investigated on SRAM's sense amplifier. Furthermore, BTI impact on other sense amplifier metrics such as sensing delay and sensing voltage are still un-explored.

Although, it is obvious that BTI can cause timing delay and reduce memory reliability in most of the designs. This paper focuses on drain input latch-type sense amplifier design due to its low power superior performance [26]. In this paper, the parameters considered for analyzing BTI impact on sense amplifier include sensing delay and sensing voltage. In this regard, the main contributions of the paper are:

- Investigation of BTI impact on the sense amplifier's sensing delay and sensing voltage.
- Analysis of BTI impact on sense amplifier synthesis with different technology nodes i.e., 90nm, 65nm, and 45nm.
- Investigation of supply voltage impact to compensate for the BTI degradation on sense amplifier.

The result depicts that as technology scales down, BTI impact on sensing delay increases, while sensing voltage decreases with a significant margin. Furthermore, the icrease in supply voltage leads to reduction in sensing delay and increase in sensing voltage, thereby leading to robust sense amplifier.

The rest of the paper is organized as follows: Section II introduces SRAM systems, drain input latch-type sense amplifier, BTI mechanism and its model. Section III gives the simulation setup, analysis metrics, and the experiments performed. Section IV analyzes the result by using different technology nodes in sense amplifier, and varying supply voltages. Finally, Section V concludes the paper

II. BACKGROUND

This section explains the functional model of an SRAM system. Afterwards, it explains the behavior of drain input latch-type sense amplifier. Finally, it presents BTI mechanism and its model analyzed in this paper.

978-1-4799-3525-3/13/$31.00 ©2013 IEEE

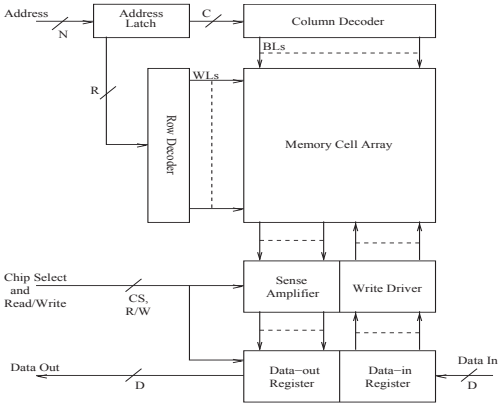


Fig. 1. Functional model of SRAM system

A. Memory model

A Memory system comprises memory cell array, row and column address decoders, read/write circuitry, input/output data registers and control logic as depicted in functional model of SRAM system in Fig. 1. [25]. However, the main focus of the paper is SRAM sense amplifier which is responsible for the amplification of small difference in the input signals.

SRAM Sense Amplifier

A sense amplifier (SA) in SRAMs takes a small voltage difference at the input (i.e., BL and BLBar) shown in Fig. 2., and produce amplified signals on the output (i.e., out and $\overline{out}$ towards memory output). There are different implementation of sense amplifier such as: strobed or non-strobed, voltage-mode or current-mode and traditional, compensated or calibrated sense amplifiers. Furthermore, there are several types of strobed voltage-mode sense amplifier such as: drain-input latch-type, gate-input stacked-latch type, look-ahead, self-closing, pulsed current source and double-tail latch-type strobed voltage-mode SAs. In this paper, drain-input latch-type SRAM sense amplifier will be focused. This design selection is based on the following advantages: (a) the design does not draw static current, (b) easily employ positive feedback to provide fast regeneration, (c) their design is straightforward, (d) the energy consumption in charging and discharging these large capacitances is reduced, and (e) the time the cell requires to develop its swing is also reduced.

The structure of the Drain input latch-type Sense Amplifier as depicted in Fig. 2. [26] consists pull-up transistors (i.e., Mup and MupBar), pull-down transistors (i.e., Mdown and Mdownbar), two access transistors (i.e., Mpass and MpassBar), two switching current source transistors (i.e., Mtop and Mbottom), and two inverters at the output of the Sense Amplifier with a load capacitance of 1fF each. The pull-up transistors and pull-down transistors are made of two PMOS (Mup and MupBar) and two NMOS (Mdown and

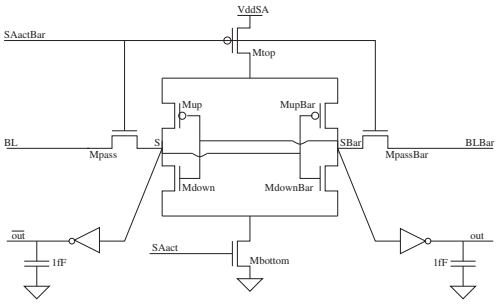


Fig. 2. Drain input latch-type Sense Amplifier

MdownBar) transistors respectively, access transistors are two NMOS devices, switching current source transistors are one PMOS and one NMOS device at the top and bottom nodes of the Drain input latch-type Sense amplifier. These explained transistors/devices receive their inputs through BL and BLbar input signals.

The operation of the sense amplifier circuitry is explained in three phases as follows: In the first phase, if SAact is low and SAactBar is high, the access transistors Mpass and MpassBar connect the BL(bar) inputs with the internal nodes S(Bar). However, when the BL input signal is pulled low, then there is a corresponding rise in the internal node SBar. However, in this phase Mtop and Mbottom transistors are in the off state. In the second phase, if SAact is high and SAactBar is low, then the pass transistors disconnect the inputs from the internal nodes. Then, node S connects MupBar and SBar connects Mup, this causes MupBar to be ON as Mtop pulls up the top node. This is the source for Mup and MupBar causing current to conduct. Furthermore, MupBar draws more current than the Mup which leads to SBar increasing faster than S, thereby leading to quick amplification of the input difference. Moreover, this is based on the PMOS transistors acting as differential amplifier with positive feedback phenomenon. Then, in the third phase, S node is actively pulled down when SBar exceeds the threshold of Mdown. All current paths are disabled when S is at 0V and SBar is at $V\mathrm{dd_{SA}}$. Next, out goes quickly down when SBar threshold exceeds the output inverter, then $\overline{out}$ is the reverse. The process is repeated when SAact an SAactBar are restored to their original values.

B. Bias Temperature Instability

Bias Temperature Instability (BTI) mechanism takes place inside MOS transistors and causes a threshold shift that translates to additional delay, as described below.

BTI Mechanism

BTI causes threshold voltage ($\mathrm{V_{th}}$) increment to MOS transistors. The $\mathrm{V_{th}}$ increment in a PMOS transistor that occurs under *negative* gate stress is referred to as NBTI, and the one that occur in an NMOS transistor under *positive* gate stress is known as PBTI. For a MOS transistor, there are

two BTI phases, i.e., the stress phase and the relaxation phase.

In recent times, exhaustive efforts has been put to understand NBTI [5]–[7], [22]. Kaczer et al. in [6] have analyzed NBTI reasonably well but have not extended their analysis to deal with NBTI at a higher level. Alam et al. in [5] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion process. The model is usable at a higher level such as circuit level. Since in this work, BTI analysis is done at the circuit level, model of [5] will be used.

Stress Phase: In the stress phase, the Silicon Hydrogen bonds (≡Si-H) break at Silicon-Oxide interface. The broken Silicon bonds (≡Si-) remain at the interface (known as interface traps), and the released H atoms/molecules diffuse towards the poly gate. The number of interface traps (N_{IT}) generated after applying a stress of time (t) is given by [5]:

$$N_{IT}(t) = \left(\frac{N_o.k_f}{k_r}\right)^{2/3} . \left(\frac{k_H}{k_{H_2}}\right)^{1/3} .(6.D_{H_2}.t)^{1/6}, \quad (1)$$

where N_o, k_f, k_r, k_H, and k_{H_2}, represent initial ≡Si-H density, ≡Si-H breaking rate, ≡Si- recovery rate, H to H_2 conversion rate, and H_2 to H conversion rate inside the oxide layer, respectively. While D_{H_2} is the hydrogen diffusion constant.

Relaxation Phase: In the relaxation phase, there is no ≡Si-H breaking. However, the H atoms/molecules diffuse back towards the interface and anneal the ≡Si- bonds. The number of interface traps that *do not* anneal by the approaching H atoms during the relaxation phase is given by [18]:

$$N_{IT}(t_o + t_r) = \frac{N_{IT}(t_o)}{1 + \sqrt{\frac{\xi.t_r}{t_o+t_r}}} \quad (2)$$

where $N_{IT}(t_o)$ is the number of interface traps at the start of the relaxation, ξ is a relaxation coefficient with ξ=0.5 [18], t_o is the duration of the previous stress phase and t_r is the relaxation duration.

Threshold voltage increment: The N_{IT} oppose the gate stress resulting in the threshold voltage increment (ΔV_{th}). The relation between N_{IT} and ΔV_{th} is given by [4]:

$$\Delta V_{th} = (1+m).q.N_{IT}/C_{ox}.\chi.\gamma, \quad (3)$$

where m, q, and C_{ox} are the holes/mobility degradation that contribute to the V_{th} increment [16], electron charge, and oxide capacitance, respectively. χ is a BTI coefficient with a value χ=1 for NBTI and χ=0.5 for PBTI [14]. Additionally, γ represents the stress duration with respect to the total input period (i.e., *activity factor*) of the transistor. The γ dependence of the ΔV_{th} shows that transistors in a gate/circuit that have different stress and relaxation phases will suffer from different degradations.

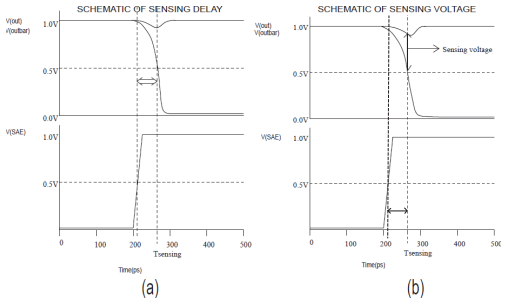


Fig. 3. Metric diagram of (a) Sensing delay and (b) Sensing voltage.

Delay increment: BTI induced ΔV_{th} of each individual MOS transistor has its contribution to the additional delay. A generalized formula that relates BTI induced ΔV_{th} in a transistor to dataline/output signal delay is given by [4], [15]:

$$\Delta D = \frac{n.\Delta V_{th}}{(V_{gs} - V_{th})} \quad (4)$$

where n is the velocity saturation index of majority carriers in MOS channels. Since NBTI causes ΔV_{th} to PMOS transistor and PBTI causes ΔV_{th} to NMOS transistor, the paper considers the threshold voltage shifts to both types of MOS transistors.

III. ANALYSIS FRAMEWORK

In this section describes the simulation setup, the metrics utilized in the experiments, and the experiments conducted in the paper.

A. Simulation Setup

A netlist of drain input latch type sense amplifier depicted in Fig. 2 has been synthesized using different technology nodes such as 90nm, 65nm, and 45nm PTM transistor models[1] and simulated using HSPICE. In the BTI analysis, the impact is added to each transistor with Verilog-A modules. Furthermore, each module generates voltage shift increment which is a function of the activity factor of the transistor.

B. Analysis Metrics

In this section, the metrics for analyzing BTI impact on sense amplifier are described. First, sensing delay, and then sensing volatge are introduced.

Sensing delay: Sensing delay metric is determined when the trigger signal (i.e., sense amplifier enable input signal) reaches 50% of the supply voltage and the target (i.e., either out or $\overline{out}$ falling output signal) reaches 50% of the supply voltage. The difference between the target and the trigger results in sensing delay as shown in Fig. 3(a). Furthermore, the relative variation of the sensing delay due to BTI is the difference between the measured sensing delay when BTI is added and referenced sensing delay when BTI is not added.

Sensing voltage: Sensing voltage metric refers to the difference between the output signals, for instance v(out) minus v($\overline{out}$) as shown in Fig. 3(b). at a fixed time (i.e., Tsensing). Tsensing is initially determined as the time when the inverted trigger signal i.e., SAE reaches 50% of the supply voltage and one of the output signals (i.e., either out or $\overline{out}$) falling output signal reaches 50% of the supply voltage in the absence of BTI. Furthermore, the relative % sensing voltage is the difference between the measured sensing voltage and the referenced sensing voltage divided by referenced sensing voltage multiplied by 100.

C. Experiments Performed

In this paper, three sets of experiments are performed to analyze BTI impacts. Initially, it presents temporal degradation of sense amplifier parameters with time. Then, it analyzes variation of BTI impact in different technology nodes. Finally, it demonstrates variations of the impact with supply voltage. These experiments are described below:

1. BTI Impact experiments: BTI impact on sensing delay and sensing voltage of the SRAM sense amplifier is investigated.

2. Technology dependent experiments: BTI impact on the sensing delay and sensing voltage of the SRAM sense amplifier synthesized from different technology nodes is investigated.

3. Supply voltage dependent experiments: BTI impact on sensing delay and sensing voltage of the SRAM sense amplifier for a particular technology node as the supply voltage increases in ascending order (i.e., from 90% of Vdd to Vdd and 110% of Vdd) is investigated as well.

IV. EXPERIMENTAL RESULTS

In this section, we present the analysis results of the experiments mentioned in the previous section.

A. Temporal BTI Impact

Initially, BTI impact on output sensing delay signals is presented and thereafter, BTI impact on output sensing voltage signals is covered.

BTI Impact on sensing delay

BTI in MOS transistors of the above mentioned SRAM sense amplifier affect activation of the amplified output signals (i.e., output signal pulled down to zero) with respect to the differential input. Differential inputs are fixed, choosen in such a manner to obtain a required output signal amplification that meets the critical timing. The differential input depends on the technology node. For example, implementing the above sense amplifier circuitry in 90nm the differential input signals is 135mv, the supply voltage is 1v, and the critical timing for which the input signal fall low is 200ps. Analysis results to depict BTI impact on sense amplifiers sensing delay is given in Fig. 4. The Figure shows that sensing delay as a function of time (i.e., aging lifetime) increases as the sense amplifier output signals ages. For this case, sensing delay increases from 30.03ps to 36.19ps which is about 20.51% increase due to BTI impact.

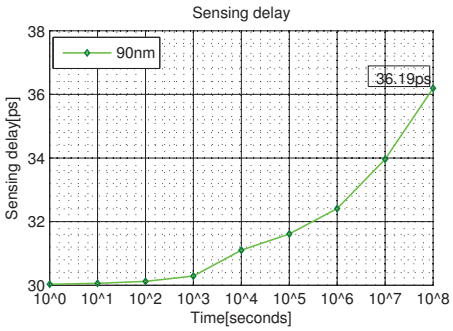


Fig. 4. BTI impact on Sensing delay for 90nm technology.

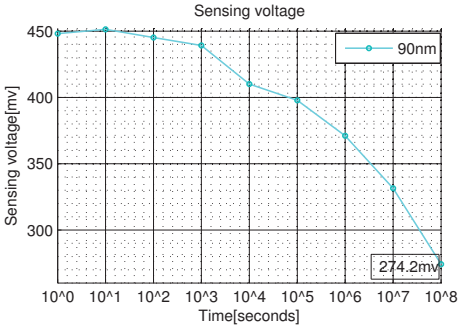


Fig. 5. BTI impact on Sensing voltage for 90nm technology.

BTI Impact on sensing voltage

In this subsection, we explore BTI impact on the sensing voltage metric considered in this paper. For this metric, the difference between two output signals (i.e., vout and v$\overline{out}$) at a given time is considered. The difference between the signals at the time of sensing changes with aging. For a given technology node (i.e., 90nm), Fig. 5 plots sensing voltage variation as a function of time (i.e., aging lifetime). This figure depicts that sensing voltage decreases with respect to increment in time. For this particular case, sensing voltage decreases from 448.1mv to 274.2mv with respect to the increase in time from 10^0 second to 10^8 seconds, and this is about -38.81% reduction in sensing voltage due to BTI impact.

B. Technology dependent BTI impact

Technology scaling down leads to oxide field increment, and this accelerates bond breaking phenomenon of BTI impact. In this section, experiments are performed for SA with different technology nodes at a nominal voltage (i.e., 0.833X of the standard supply voltage) to achieve the required impact [26]. For example, standard supply voltage of 45nm node is 1.0v and when multiplied by this factor it results to 0.833v. However, experiments are performed on different technology

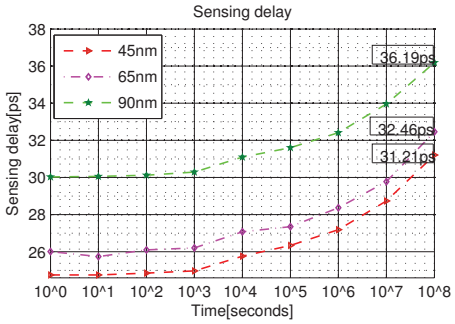


Fig. 6. Technology dependent Sensing delay.

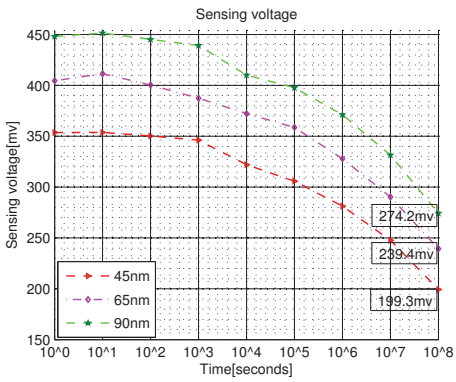


Fig. 7. Technology dependent Sensing voltage.

nodes as 45nm, 65nm, and 90nm. Experiments are performed to explore first, BTI induced sensing delay increment in different technologies. Thereafter, the sensing voltage technology dependence is investigated.

Sensing delay variation

Sensing delay (i.e., interval between SAE activation and falling output signal as shown in Fig. 3(a)) is strongly dependent on technology nodes. For instance, as shown in Fig. 6., in degradation free case, 90nm, 65nm, and 45nm based sense amplifier have the sensing delay of 30.03ps, 26.01ps, and 24.75ps, respectively. The figure also shows BTI induced degradation in sensing delay of sense amplifier based on different technology nodes i.e., 31.21ps, 32.46ps, and 36.19ps in 45nm, 65nm, and 90nm, respectively. There is significant variation in the increment. For instance, the variation is 20.51% in 90nm, and approaches 24.80%, and 26.10% in 65nm, and 45nm, respectively. An important point in the result is that although the impact on 45nm is higher with a small margin. This shows that the smaller the technology node, the higher the impact on sensing delay.

Sensing voltage variation

Sensing voltage (i.e., voltage difference between the bit lines at a particular instant of time as shown in Fig. 3(b)) varies significantly with technology scaling. Fig. 7. represent reduction in the sensing voltage with technology scaling. For instance, in degradation free case, the sensing voltage is 448.1mv for 90nm and approaches 404.6mv and 353.6mv in 65nm and 45nm technology base sense amplifier, respectively. The figure also shows BTI induced degradation in sensing voltage of SA based on different technology i.e., 199.3mv, 239.4mv, and 274.2mv in 45nm, 65nm, and 90nm, respectively. There is significant variation in the reduction. For instance, the variation is -38.81% in 90nm and approaches -40.83% and -43.67% in 65nm and 45nm, respectively. This also shows that the lower the technology node, the higher the impact on sensing voltage.

C. Supply voltage dependent BTI impact

This section presents result of supply voltage dependence experiments. First, supply voltage dependence of sensing delay is investigated. Afterwards, supply voltage dependence of sensing voltage is explored.

Sensing delay variation

Transistors experience significant supply voltage variation during operation. The variation affect oxide field and consequently BTI impact. To investigate the variation, the supply voltage is +/-10 of its nominal Vdd. Fig. 8. shows the investigation results. The figure shows that for the degradation free case, by lowering the supply voltage, sensing delay increases significantly. Additionally, impact of BTI is more significant at lower supply voltage. Fig. 8. represent increment in sensing delay with supply voltage reduction. For instance, in degradation free case, the sensing delay is 22.44ps for 0.9163v and approaches 24.75ps and 26.10ps for 0.8330v and 0.7497v supply voltage, respectively. The figure also shows

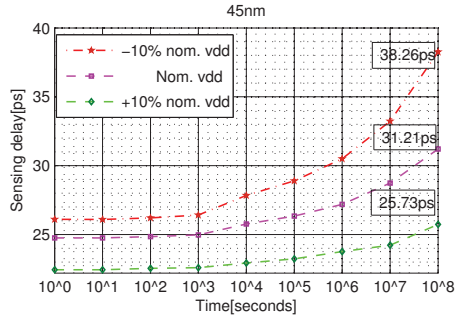


Fig. 8. 45nm supply voltage dependent sensing delay.

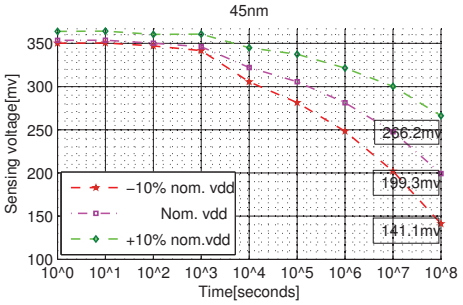


Fig. 9. 45nm supply voltage dependent sensing voltage.

BTI induced degradation in sensing delay of sense amplifer based on different supply voltages, i.e., 25.73ps, 31.21ps, and 38.26ps for 0.9163v, 0.8330v, and 0.7497v respectively. This shows % increment of 14.66% for 0.9163v and 26.10% and 46.59% for 0.8330v and 0.7497v respectively. This shows that higher supply voltage reduces the sensing delay with approx. 18% from the nominal voltage.

Sensing voltage variation

In this section, supply voltage dependence of sensing voltage is presented. Supply voltage varies significantly with sensing voltage. To investigate the variation, Fig. 9. shows the results. The figure shows that for the degradation free case by increasing the supply voltage, sensing voltage increases significantly. Besides, impact of BTI is more significant at higher supply voltage. Fig. 9. represent increment in sensing voltage with supply voltage increment. For instance, in degradation free case, sensing voltage is 350.2mv for 0.7497v and approaches 353.6mv and 363.9mv for 0.8330v and 0.9163v supply voltage, respectively. The figure also shows BTI induced degradation in sensing voltage of sense amplifer based on different supply voltages, i.e., 141.1mv for 0.7497v and approaches 199.3mv and 266.2mv for 0.8330v and 0.9163v, respectively. This shows % reduction of sensing voltage is -26.85% for 0.9163v and -43.64% and -59.71% for 0.833v and 0.7497v respectively. This implies that the higher supply voltage, BTI degradation with 33% for sensing voltage from the nominal supply voltage.

V. CONCLUSION

This paper investigates the impact of Bias Temperature Instability (BTI) on drain input latch-type memory sense amplifier of different technologies. First, BTI impact increases the sensing delay and reduces the sensing voltage causing the memory sense amplifier to be less reliable and robust. Second, increase in supply voltage per technology node compensate the sensing delay and sensing voltage causing the sense amplifier to be more robust and reliable. These results are validated with HSPICE.

REFERENCES

[1] ITRS, “International Technology Roadmap for Semiconductor 2004” ”www.itrs.net/common/2004 update/2004update.htm.”.
[2] S. Borkar, et al “Micro architecture and Design Challenges for Giga scale Integration”, *Pro. of Intl. Sympos. Micro architecture*, 2004.
[3] S. Hamdioui et al, “Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes”, *DATE*, 2013.
[4] B.C. Paul, K. Kang, H. Kufluoglu, M.A. Alam, K. Roy, “Impact of NBTI on the Temporal Performance Degradation of Digital Circuits”, *IEEE Electron Device Letter*, Vol. 26, No.8, Aug. 2005.
[5] M. A. Alam and S. Mahapatra, “A Comprehensive Model of PMOS NBTI Degradation”, *Microelectronics Reliability* , Vol:45, 2005.
[6] B. Kackzar, et al., “Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification”, *Proc. of Intl. Physics Reliability Symp.(IPRS)*, pp: 381-387, 2005.
[7] S. Zafar, Y.H. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, M. Chudzik, “A comparative study of NBTI and PBTI in SiO2/HfO2 stacks with FUSI, TiN gates”, *Pro. of VLSI Technology symp.*, 2006.
[8] N. Kizmuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, T. Horiuchi, “The Impact of BTI for Direct Tunneling Ultra Thin Gate Oxide of MOSFET Scaling”, *VLSI Technology, Digest of Technical Papers.*, pp: 73-74, 1999.
[9] P. Pouyan et al, “Process Variability-Aware Proactive Reconfiguration Technique for Mitigating Aging effects in Nano Scale SRAM lifetime”, *IEEE 30th VLSI Test Symposium.*, 2012.
[10] S Hamdioui, ”Testing Embedded Memories: A Survey”, *Mathematical and Engineering Methods in Computer Science*, pp. 32-42, 2013.
[11] S Hamdioui, R Wadsworth, JD Reyes, AJ Van De Goor, ”Memory Fault Modeling Trends: A Case Study”, *Journal of Electronic Testing 20 (3)*, pp. 245-255, 2004.
[12] L. Dilillo, “Resistive-Open Defects in Embedded-SRAM core cells: Analysis and March Test Solution”, *Proc. of ATS*, pp: 266-271, 2004.
[13] D. Rodopoulos, S.B. Mahato, V.V. de Almeida Camargo, B. Kaczer, F. Catthoor, S. Cosemans, G. Groeseneken, A. Papanikolaou, D. Soudris, “Time and Workload Dependent Device Variability in Circuit Simulations“ *Proc. Intl. Conf on IC Design and Technology*, pp: 1-4, 2011.
[14] M. T. Luque, B. Kaczer, J. Franco, P.J. Roussel, T. Grasser, T.Y. Hoffmann, G. Groeseneken, “From Mean Values to Distribution of BTI Lifetime of Deeply scaled FETs through Atomistic Understanding of the Degradation“ *Sym. on VLSI Technology*, pp: 152-153, 2011.
[15] T. Sakurai, and A.R. Newton, “Alpha-Power law MOSFET model and its applications to CMOS delay and other formulas”, *IEEE J. Solid-State Circuits*, Vol.25, No.2, April 1990,
[16] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, S. Krishnan, “NBTI impact on transistor and circuit: Models, mechanisms and scaling effects”, *Pro. of IEDM*, 2003.
[17] B. Cheng, A. R. Brown, “Impact of NBTI/PBTI on SRAM Stability Degradation”, *IEEE ELECTRON DEVICES LETTERS*, 2011.
[18] S. Kumar, C.H. Kim, S. Sapatnekar, “Impact of NBTI on SRAM Read Stability and Design for Reliability”, *Pro. of ISQED*, pp: 212-128, 2006.
[19] A. Bansal et al, “Impact of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability”, *J. Microelectronics Reliability*, 2009.
[20] S. Hamdioui, et al., ‘An experimental analysis of spot defects in SRAMs: realistic fault models and tests”, *Proc. of 9th Asian Test Symp.*. pp: 131 - 138 , 2000
[21] S. Hamdioui, Z. Alars and A.J. van de Goor, “Opens and Delay Faults in CMOS circuits”, *IEEE Transactions on Computers*, Dec. 2006.
[22] S. Khan, M. Taouil, S. Hamdioui, H. Kukner, P. Raghavan, F. Catthoor, “Impact of Partial Resistive Defects and Bias Temperature Defects and Bias Temperature Instability on SRAM Decoder Reliability”, *Pro. of 7th IEEE International Design and Test Symposium*, 2012.
[23] R. Menchaca, H. Mahmoodi, “Impact of Transistor Aging Effects on Sense Amplifier Reliability in Nano-Scale CMOS”, *13th Int’l Sym. on Quality Electronic Design*, 2012.
[24] V. Chandra, R. Aitken, “Impact of Voltage Scaling on Nanoscale SRAM Reliability”, *DATE*, 2009.
[25] H. Kukner, “Generic and Orthogonal March Element based Memory BIST Engine”, *MSc Thesis TU Delft, The Netherlands*, 2011.
[26] S. Cosemans, “Variability-aware design of low power SRAM memories”, *Ph.D Thesis Katholieke Universiteit Leuven*, 2009.
[27] Predictive Technology Model ”http://ptm.asu.edu/”,

Integral Impact of BTI and Voltage Temperature Variation on SRAM Sense Amplifier

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Halil Kukner Pieter Weckx Praveen Raghavan Francky Catthoor
IMEC vzw
Kapeldreef 75, 3001 Leuven, Belgium
{Halil, Pieter.Weckx, Ragha, Francky.catthoor}@imec.be

***Abstract*—With the continuous downscaling of CMOS technologies, ICs become more vulnerable to transistor aging mainly due to Bias Temperature Instability (BTI). A lot of work is published on the impact of BTI in SRAMs; however most of the work focused mainly on the memory cell array. An SRAM consists also of peripheral circuitries such as address decoders, sense amplifiers, etc. This paper characterizes the combined impact of BTI and voltage temperature fluctuations on the memory sense amplifier for different technology nodes (45nm up to 16nm). The evaluation metric, the sensing delay (SD), is analyzed for various workloads. In contrast to earlier work, this paper thoroughly quantifies the increased impact of BTI in such sense amplifiers for all the relevant technology scaling parameters. The results show that the BTI impact for nominal voltage and temperature is 6.7% for 45nm and 12.0% for 16nm when applying the worst case workload, while this is 1.8% for 45nm technology and 3.6% higher for 16nm when applying the best case workload. In addition, the results show that the increase in power supply significantly reduces the BTI degradation; e.g., the degradation at $-10\%V_{dd}$ is 9.0%, while this does not exceed 5.3% at $+10\%V_{dd}$ at room temperature. Moreover, the results that the increase in temperature can double the degradation; for instance, the degradation at room temperature and nominal V_{dd} is 6.7% while this goes up to 18.5% at 398K.**

***Index Terms*—BTI, NBTI, PBTI, SRAM sense amplifier**

I. INTRODUCTION

In recent decades, CMOS technology has been sustained with aggressive downscaling that severely impacts the reliability of devices [1,2,28]. These trends are due to advancements in the fabrication technology, introduction of novel materials and evolution of architecture designs. Bias Temperature Instability (BTI) (i.e., Negative BTI in PMOS transistors and Positive BTI in NMOS transistors) is a reliability failure mechanism which affects the performance of MOS transistors by increasing their threshold voltage and reducing their drain current (I_d) over the operational lifetime [5,11]. However, studies of such individual devices or small composites do not allow extrapolation of these effects on larger circuits like SRAMs.

Static Random-Access Memories (SRAM) occupy a large fraction of semiconductor chip and play a major role in the silicon area, performance, and critical robustness [12]. An SRAM system consists of an array of cells, its peripherals circuits such as row and column address decoders, control circuits, write drivers, and sense amplifiers. Designing an optimal reliable memory system requires the consideration of all its sub-parts, i.e., their degradation rate depends on the application (e.g. workload, temperature, etc); for instance, the aging rate of the sense amplifier may differ from that of the memory array and other peripheral circuits.

Many publications analyzed the BTI impact on SRAM cell array, while very limited work is published on the SRAM peripheral circuitry. For instance, Binjie *et al.* [17] investigated NBTI impact on Static Noise Margin (SNM) and Write Noise Margin (WNM) degradation of 6T SRAM cell. Kumar *et al.* [18] Analyzed the impact of NBTI on the read stability and SNM of SRAM cells. Andrew [19] investigated the mechanism of NBTI degradation on SRAM metrics such as SNM. Bansal et al [20] presented insights on the stability of an SRAM cell under the worst-case conditions and analyzed the effect of NBTI and PBTI, individually and collectively. Rodopoulos *et al.* [22] investigated the atomistic pseudo-transient BTI simulation with built-in workloads. Khan et al [23] investigated BTI analysis of FinFET based SRAM cell.

On the other hand, few authors have focused on reliability analysis of the SRAM peripheral circuit. Khan *et al.* [21] investigated the impact of partial opens conjuction BTI in SRAM address decoders. Menchaca *et al.* [24] analyzed the BTI impact on different sense amplifier designs implemented on 32nm technology node by using failure probability (i.e., flipping a wrong value) as a reliability metric. Agbo *et al.* [25] investigated the BTI impact on SRAM drain-input latch type sense amplifier design implemented on 90nm, 65nm, and 45nm for different supply voltages by using sensing delay and sensing voltage as reliability metrics. However, quantitative analysis of BTI impact of peripheral circuits (including sense amplifiers) while considering different workloads, temperatures, supply voltages and how they correlate with technology scaling is still to be explored. It is worth noting that understanding and quantifying the aging rate of each memory part is needed for optimal reliable memory design; this is because the different parts may degrade with different rates depending e.g. on the workload (application).

This paper focuses on standard latch-type sense amplifier design due to its superior performance [30] and analyzes the BTI impact for different temperatures and supply voltages, and different workloads. The main contributions of the paper are as follows:

- Investigation of BTI impact on the sense amplifier's sensing delay. In contrast to previous work, we analyze the BTI stress and relaxation cycles for each transistor individually to obtain more accurate results. These stress and relaxation cycles are workload dependent. In this

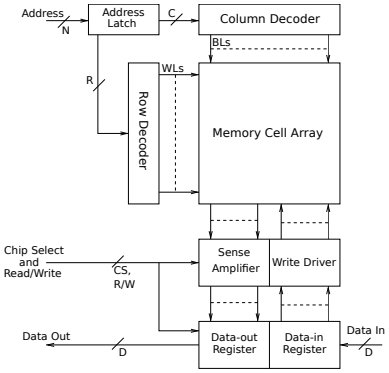


Fig. 1. Functional model of SRAM system

work we define eight realistic workloads.

- Investigation of BTI impact on different workloads for different technology nodes.
- Thorough quantitative analysis of the BTI impact on the sense amplifier for deeply nano-scaled technology nodes, i.e., 45nm, 32nm, 22nm, and 16nm.
- Analysis of BTI impact under different supply voltages and temperatures on the SRAM sense amplifier sensing delay using different workloads.

The rest of the paper is organized as follows: Section II introduces the SRAM model, standard latch-type sense amplifier, BTI mechanism and its model. Section III provides our analysis framework, analysis metric, and the performed experiments. Section IV analyzes the result for different technology nodes, workloads, varying supply voltages and temperatures. Finally, Section V concludes the paper.

II. BACKGROUND

This section presents first the functional SRAM model. Afterwards, it focuses on the behavior of the standard latch-type sense amplifier. Finally, it explains the BTI mechanisms and its model analyzed in this paper.

A. Memory model

Figure 1 depicts a functional model of the SRAM system [28]. A memory system is comprised of a memory cell array, row and column address decoders, read/write circuitry, input/output data registers and control logic. The main focus of the paper is the sense amplifier.

SRAM Sense Amplifier

Several implementations of sense amplifiers have been proposed. In this paper, the standard latch-type SRAM strobed sense amplifier will be addressed which is representative for industrial SA designs [30].

The structure of the Standard latch-type Sense Amplifier is depicted in Fig. 2. The width length ratio of each transistor is presented by W/L.

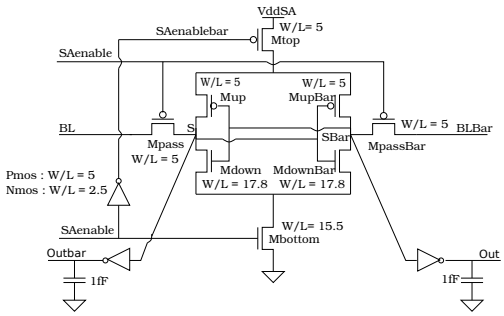


Fig. 2. Standard latch-type Sense Amplifier

The operation of the sense amplifier consists of two phases. In the first phase, when SAenable is low, the access transistors Mpass and MpassBar connect to the BL (BLBar) with the internal nodes S (SBar). In this phase, Mtop and Mbottom transistors are switched off. In the second phase, when SAenable is high, the pass transistors disconnect the BL (BLBar) input from the internal nodes. The cross coupled inverters get their current from Mtop and Mbottom and subsequently amplify the difference between S and SBar and produce digital outputs on Out and Outbar. S (SBar) node is actively pulled down when SBar (S) exceeds the threshold voltage of Mdown. The positive feedback loop ensures low amplification time and produces the read value at its output. Moreover, all current paths are disabled when S (SBar) is at 0V and SBar (S) is at V_{ddSA} or vice versa. This process is repeated for each read operation.

B. Bias Temperature Instability

BTI mechanism takes place inside the MOS transistors and causes a threshold voltage shift that impacts the delay negatively; its mechanism is described below.

BTI Mechanism

BTI increases the absolute V_{th} value in MOS transistors. For PMOS, negative BTI (NBTI) reduces the V_{th} while for NMOS, positive BTI (PBTI) the V_{th} increases. Recently, exhaustive efforts have been put to understand NBTI [5,10,11]. Kaczer *et al.* in [10] have analyzed NBTI using an atomistic model. Alam *et al.* [5] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion process. The model is usable at a higher level such as circuit level. In Kukner *et al.* in [26], both the atomistic and RD models have been compared. The authors conclude that for the long-term simulation time, RD model is lightweight than the atomistic model. For this reason we select this model [5]. For a MOS transistor, there are two BTI phases, i.e., the stress phase and the relaxation phase.

Stress Phase: In the stress phase, the Silicon Hydrogen bonds (≡Si-H) break at Silicon-Oxide interface. The broken Silicon bonds (≡Si-) remain at the interface (known as interface traps), and the released H atoms/molecules diffuse towards the gate

oxide. The number of interface traps (N_{IT}) generated after applying a stress of time (t) is given by [5]:

$$N_{IT}(t) = \left(\frac{N_o \cdot k_f}{k_r}\right)^{2/3} \cdot \left(\frac{k_H}{k_{H_2}}\right)^{1/3} \cdot (6 \cdot D_0 \cdot t)^{1/6} \quad (1)$$

where N_o, k_f, k_r, k_H, and k_{H_2}, represent initial ≡Si-H density, ≡Si-H breaking rate, ≡Si- recovery rate, H to H_2 conversion rate, and H_2 to H conversion rate inside the oxide layer, respectively. D_0=D_{H_2}.exp($-E_A$/kT) [6] is the diffusion coefficient of the produced H_2 species and E_A is the activation energy, k is the boltzman constant, and T is the temperature in Kelvin.

Relaxation Phase: In the relaxation phase, there is no ≡Si-H breaking. However, the H atoms/molecules diffuse back towards the interface and anneal the ≡Si- bonds. The number of interface traps that *do not* anneal by the approaching H atoms during the relaxation phase is given by [18]:

$$N_{IT}(t_o + t_r) = \frac{N_{IT}(t_o)}{1 + \sqrt{\frac{\xi \cdot t_r}{t_o + t_r}}} \quad (2)$$

where $N_{IT}(t_o)$ is the number of interface traps at the start of the relaxation, ξ is a relaxation coefficient with ξ=0.5 [18], t_o is the duration of the previous stress phase and t_r is the relaxation duration.

Threshold voltage increment: The N_{IT} oppose the gate voltage which result in a threshold voltage increment (ΔV_{th}). The relation between N_{IT} and ΔV_{th} is given by [4]:

$$\Delta V_{th} = (1 + m) \cdot q \cdot N_{IT}/C_{ox} \cdot \chi, \quad (3)$$

where m, q, and C_{ox} are the holes/mobility degradation that contribute to the V_{th} increment [16], electron charge, and oxide capacitance, respectively. χ is a BTI coefficient with a value χ=1 for NBTI and χ=0.5 for PBTI [14].

III. ANALYSIS FRAMEWORK

In this section, the analysis framework of the standard latch type sense amplifier circuit is described. Furthermore, the workloads used in performing the experiment are explained. Thereafter, the output performance metric is presented. Finally, the conducted experiments are presented.

A. Framework Flow

Figure 3 depicts a flexible and generic BTI framework for the standard latch type sense amplifier circuit. The framework evaluates the BTI impact for different designs, technologies, workload under normal conditions and considering VT (i.e., voltage, and temperature) variations. The framework consist of a MATLAB and HSPICE working environment. The MATLAB environment typically is used for pre-processing and post-processing, it prepares BTI augmented files to run in HSPICE. The results of HSPICE that simulates the BTI augmented netlist, are subsequently post-processed in MATLAB. Furthermore, the framework analysis is divided into three parts (i.e., input, processing, and output blocks) and they are explained next.

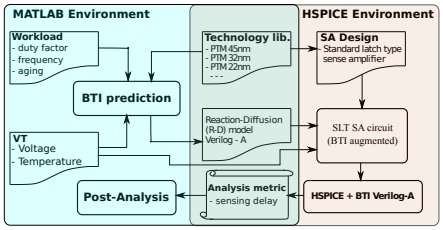


Fig. 3. Analysis framework for the standard latch sense amplifier circuit.

Input: The general input blocks of the framework are the SA design, technology library, workload, and voltage temperature (VT). They are explained as follows.

- SA design: Generally, all sense amplifier design can be used. In this paper we focus only on the standard latch type sense amplifier. The SA design is described by an HSPICE netlist.
- Technology library: Different technology nodes are considered in this work, they are 45nm, 32nm, 22nm, and 16nm and are obtained from PTM library cards [31].
- VT: This block specifies the temperatures and voltages. In this paper, we restrict ourselves to temperatures T_1 = 298K, T_2 = 348K, and T_3 = 398K and supply voltages V_1 = -10%V_{dd}, V_2 = V_{dd}, and V_3 = +10%V_{dd}. Note that each technology has its own nominal voltage.
- Workload: The shift in threshold voltage is a function of stress and relaxation durations of the transistors. This implies that BTI degradation depends on the amount of ON and OFF (idle) states of the input patterns which translates to workload. To perform this analysis, we assume that today's application consists of 10% - 90% memory instructions and the percentage of read instructions is typically 50% - 90%. Furthermore, we derive from this the following cases: best case with stress period of 0.1 * 0.1 = 0.01, worst case with 0.9 * 0.9 = 0.81, and mid case: 0.5 * 0.5 = 0.25. They lead to the following workload sequences: $S1$: $R0I^{99}$, $S2$: $R0R1I^{198}$, $S3$: $R0^4I^1$, $S4$: $(R0R1)^4I^2$, $S5$: $(R0)I^{24}$, $S6$: $(R0R1)I^{24}$, $S7$: $(R0)I^{50}$, and $S8$: $(R0R1)I^{50}$. In these sequences, $R0$ stands for read 0, $R1$ stands for read 1, I for idle operation (which includes memory write operations). For example, $S1$: $R0I^{99}$ is workload where read 0 is followed by 99 idle operations. The best and worst case will be analyzed in most detail.

The workload inputs are typically characterized by their duty factor, frequency, and aging (or stress time). The BTI impact sensitivity is highly dependent on the input stimulus clock cycle (i.e., frequency), its aging and duty factor (DC stress or AC stress) w.r.t., the affected device or circuitry.

i **Frequency** The BTI induced degradation depends on the signal frequency to the sense amplifier design. In this experiment, the frequencies considered for the SA design are 1.32GHz, 1.89GHz, 2.70GHz, and 3.86GHz for 45nm,

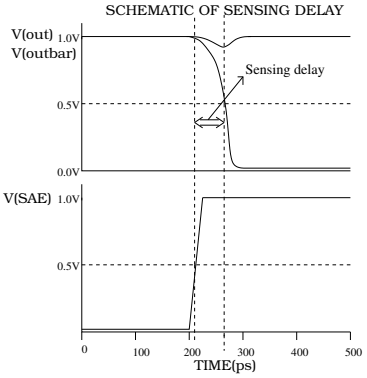


Fig. 4. Metric diagram of Sensing delay.

32nm, 22nm, and 16nm technology nodes, respectively.

ii **Aging** The BTI induced degradation depends strongly on the stress time. The stress time defines how long the workload sequence is being applied. A workload sequence is assumed to be repeated until the age time is reached.

iii **Duty Factor** The input signal is a function of the duty factor that affects the BTI induced degradation of the sense amplifier design. The duty factor of the input SA enable signal (see Figure 4.) is approximately 0.48. This applies only to read operations. During write or idle operations, SAenable signal is disabled. From the waveform analysis, we extract for all transistors individually their stress and relaxation cycles; thereby is able to obtain accurate duty cycles for each transistor. This enhances the accuracy of our simulation results. Based on the duty factor and age time, interface traps (Eqns. 1 and 2) or threshold voltage increments (Eqn. 3) can be attributed to all transistors in an accurate manner.

Processing: There are two processing blocks, the BTI predictor and the HSPICE simulation unit. The long term BTI predictor uses the duty factor, frequency, and aging to predict the interface traps/ threshold voltage increment of each device in the sense amplifier. In addition to workload inputs, inputs are required from the reaction-diffusion model (such as K_f, K_r, D_H, etc.), technology parameters, and voltage temperature (VT). Once the BTI induced V_{TH} increments are calculated per transistor, the original BTI free netlist will be updated. This new netlist is simulated in HSPICE/Verilog-A.
Output: Finally, post-analysis of the results are performed for varying voltages and temperatures in MATLAB environment.

B. Output Analysis Metrics

In this section, the sensing delay metric used for analyzing BTI impact on sense amplifier is described.

Sensing delay: The sensing delay metric is determined when the trigger signal (i.e., sense amplifier enable input signal) reaches 50% of the supply voltage and the target (i.e., either out or outbar falling output signal) reaches 50% of the supply voltage. The difference between the target and the trigger

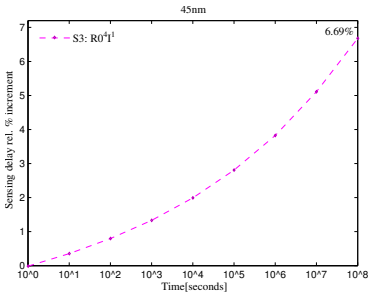


Fig. 5. BTI impact on Sensing delay.

results in sensing delay as shown in Fig. 4. Furthermore, the relative variation of the sensing delay due to BTI is the difference between the measured sensing delay when BTI is added and referenced sensing delay when BTI is not added.

C. Experiments Performed

In this paper, four sets of experiments are performed to analyze BTI impacts. These experiments are described below:

1 **BTI Impact Experiments:** BTI impact on sensing delay of the SRAM sense amplifier is investigated.
2 **Workload Dependent Experiments:** BTI impact on the sensing delay of the SRAM sense amplifier for different workloads on different technology nodes is investigated.
3 **Technology Dependent Experiments:** BTI impact on sensing delay using different technology nodes is investigated.
4 **Supply Voltage and Temperature Dependent Experiments:** BTI impact on sensing delay of the SRAM sense amplifier for varying supply voltages (i.e., $-10\%V_{dd}$, V_{dd} and $+10\%$ V_{dd}) and temperatures (i.e., $298K$, $348K$ and $398K$) for different technology nodes are explored.

IV. EXPERIMENTAL RESULTS

In this section, we present the analysis results of the experiments mentioned in the previous section.

A. Temporal BTI Impact

The BTI in MOS transistors affect the sensing delay of the sense amplifier, i.e., the time required to amplify the input from BL and BLBar to outputs Out and Outbar (see Figure 2). In order to quantify this delay, we simulate the initial BTI-free SA design, for each technology node and take their sensing delays as references. To obtain proper sensing delays, appropriate values of BL and BLBar should be selected. For 45nm, we assume the differential input to be 100mV (V_{dd}) [30]. Subsequently, we modify this differential voltage in such a way to meet up with the frequencies of the lower technology nodes. Note that the transistors are scaled with each technology.
Figure 5 shows the relative increment of the sensing delay w.r.t., the stress time (aging) for workload $S3$ using 45nm technology. The figure shows a quadratic delay increment with respect to the stress time. For example, after 10^8sec the delay increments equals 6.69% due to BTI.

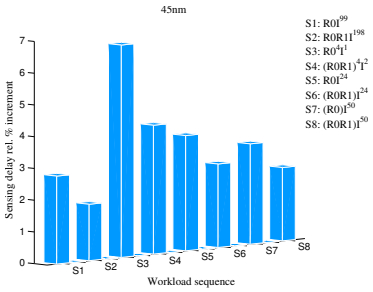


Fig. 6. Workload dependent Sensing delay.

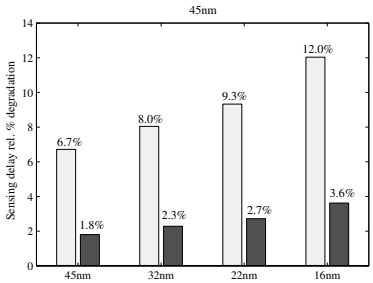


Fig. 7. BTI impact on Sensing delay for all technology nodes.

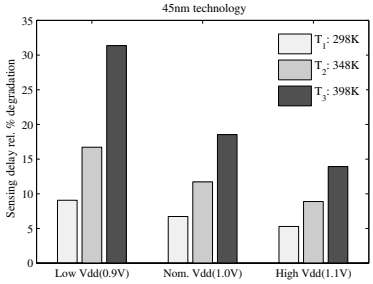


Fig. 8. Worst case sensing delay for supply voltage and temperature variations.

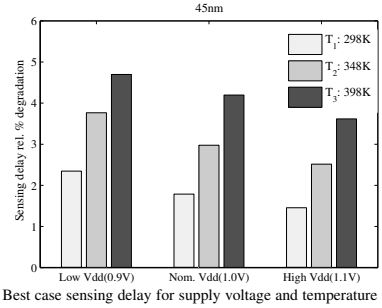


Fig. 9. Best case sensing delay for supply voltage and temperature variations.

B. Workload Dependency

The BTI induced degradation is sensitive to the workload. Hence, this workload-dependent behaviour is one of the main contributions of this study. A better understanding of this behaviour will strongly help to select the proper mitigation schemes and to reduce the cost for highly dynamic cost-sensitive embedded systems. The workload defines when and how long each transistor is stressed. Figure 6 shows the relative BTI induced sensing delay for sequences $S1$, $S2$, $S3$, $S4$, $S5$, $S6$, $S7$, and $S8$ respectively. There is a significant variation in the relative sensing delay increment. For instance, workload sequence, $S2$ ($R0R1I^{198}$) has a lower impact as it is activated (stressed) only 1% of the signal duration than the other workloads, whereas workload sequence, $S3$ ($R0^4I^1$) has the highest impact as it is activated (stressed) 80% of the signal duration. The remaining workloads result in a delay increment between the extreme cases $S2$ and $S3$. The same trends are observed for the other technology nodes.

C. Technology Dependency

CMOS technology scaling results in an apparent oxide field increment which speeds up the BTI covalent bond breaking phenomenon and thus the BTI induces threshold voltage. Therefore, it is essential to evaluate the reliability of different technology nodes. Experiments are performed at nominal supply voltage and temperature (T_1 = 298K) for different technology nodes with their corresponding supply voltages 1.0V, 0.9V, 0.8V, and 0.7V for 45nm, 32nm, 22nm, and 16nm, respectively. Here, we focus only on the best and worst case workloads $S2$ and $S3$ respectively. Experiments are performed to investigate the impact of BTI on sensing delay for different technologies for worst and best case workloads, i.e., workloads S3 and S2, respectively. Figure 6 depicts the sensing delay of these experiments for a stress time of 10^8s. The figure shows for both the worst case and best case workloads that the relative delay increment increases with advanced technology nodes. However, this increment is larger for the worst case workload. For instance, the variation for the worst case workload increases from 6.7% for 45nm to 12.0% for 16nm, while for the best case the increment is 1.8% for 45nm and 3.6% for 16nm only.

D. Supply Voltage and Temperature Dependency

Conventionally, supply voltage and temperature fluctuations impact the operating condition of MOS transistors. Supply voltage variations in a transistor can impact the sensing delay significantly as it impacts the operational speed. In addition, variation in supply voltage also affects the oxide field (capacitance) and subsequently the BTI impact (C_{ox} in Eqn. 3). The analysis of the supply voltage variation is performed on two workloads, (i.e., worst-case (S3) and best-case (S2)) for 45nm technology. The supply voltage is varied between -10% and +10% of nominal V_{dd}, i.e., between 0.9V and 1.1V. Figures 8 and 9 depict the BTI induced sensing delay for various supply voltages and temperatures for the worst-case and best-case workloads, respectively. The figures show for different voltages, i.e., low V_{dd} (Low V_{dd} = 0.9V), nominal V_{dd} (Nom. V_{dd} = 1.0V), and high V_{dd} (High V_{dd} = 0.9V) and

temperatures (i.e., T_1 = 298K, T_2 = 348K, and T_3 = 398K) the impact on the sensing delay (vertical axis). From the figures we conclude the following:

- Increasing the temperature leads to a higher BTI induced degradation. For example, in Figure 8, for a fixed supply voltage, e.g., V_{dd} = 1.1V, the BTI induce degradation is 5.3% for T_1, and while 8.9% and 13.9% for T_2 and T_3, respectively. In addition to that, the same trend is observed at V_{dd} = 1.0V and V_{dd} = 0.9V. In Figure 9, for considering V_{dd} = 1.1V as a fixed reference, the relative sensing delay degradation becomes 1.5% for T_1 while 2.5% and 3.6% for T_2 and T_3, respectively.
- Increasing the supply voltage reduces the sensing delay degradation. For instance, in Figure 8, for a fixed temperature, e.g., T_1, the BTI induced degradation on the sensing delay at 10^8s is 9.0% for low V_{dd}, while 6.7%, and 5.3% for nominal V_{dd} and high V_{dd}, respectively. Moreover, the same trend is observed at T_2 and T_3. In Figure 9, for fixed T_1, the relative sensing delay degradation is 2.3% for V_{dd} = 0.9V, while 1.8% and 1.5% for V_{dd} = 1.0V and V_{dd} = 1.1V, respectively.
- The performance degradation is much higher for the worst-case workload (S3) as compared to the best-case workload (S2). For instance, in Figure 8, at low V_{dd} and T_3 = 398K, the sensing degradation is 31.3% while in Figure 9, for the same voltage and temperature, the BTI induced degradation is only 4.7%.

In conclusion, based on the experiments, it is extremely important for designers to include proper design margins to guarantee the life-time operation of the considered SA circuitry under given operating conditions (such as voltage and temperature). In order to design a reliable memory system, designers need to understand the degradation of each sub-component. Based on such information, monitoring and mitigation schemes might be considered for SAs. For example, a sensing delay monitoring circuit could be used as sensor to adaptively control the supply voltage. Another approach could focus on redesigning more robust SA, by increasing the drive strength (or device width) of critical transistors [29].

V. CONCLUSION

This paper investigated the combined impact of Bias Temperature Instability (BTI), voltage and temperature variation and different workloads on the standard latch type memory sense amplifier for different technologies. The results show that the sensing delay degradation is strongly workload dependent and reaches up to 12.8%. Both the scaling and increase in temperature severely impact the BTI degradation. Increasing the supply voltage reduces the BTI induced degradation leading to more reliable and robust sense amplifiers, but at the cost of a higher power consumption. Optimizing a reliable memory system requires the consideration of all its sub-parts, i.e., their degradation rate depends on the application (e.g. workload, temperature, etc); for instance, the aging rate of the sense amplifier may differ from that of the memory array and other peripheral circuits.

REFERENCES

[1] ITRS, "International Technology Roadmap for Semiconductor 2004" "www.itrs.net/common/2004 update/2004update.htm.".

[2] S. Borkar, et al "Micro architecture and Design Challenges for Giga scale Integration", *Pro. of Intl. Sympos. Micro architecture*, 2004.

[3] S. Hamdioui et al, "Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes", *DATE*, 2013.

[4] B.C. Paul et al, "Impact of NBTI on the Temporal Performance Degradation of Digital Circuits", *IEEE Electron Device Letter*, Vol. 26, No.8, Aug. 2005.

[5] M. A. Alam et al, "A Comprehensive Model of PMOS NBTI Degradation", *Microelectronics Reliability* , Vol:45, 2005.

[6] D. Varghese, et al, "On the Dispersive versus Arrhenius Temperature Activation of NBTI Time Evolution in Plasma Nitrided Gate Oxides: Measurements, Theory, and Implications", *IEDM*, Dec. 2005, pp. 1-4.

[7] K. Kang, et al, "Estimation of Statistical Variation in Temporal NBTI Degradation and Its Impact on Lifetime Circuit Performance", *ICCD*, 2005, pp. 730-734.

[8] R. Wang, et al., Threshold Voltage Variation with Temperature in MOS Transistors, *IEEE Transaction on Electron Devices*, pp: 386- 388, 1971.

[9] S. Sapatnekar, et al., Overcoming Variatins in Nano-scale Technologies, IEEE Transaction on Emerging and Selected Topics in Circuits and Systems, pp: 5-18, 2011.

[10] B. Kackzar, et al., "Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification", *IPRS* , pp: 381-387, 2005.

[11] S. Zafar, et al, "A comparative study of NBTI and PBTI in SiO2/HfO2 stacks with FUSI, TiN gates", *Pro. of VLSI Technology symp.*, 2006.

[12] P. Pouyan, et al, "Process Variability-Aware Proactive Reconfiguration Technique for Mitigating Aging effects in Nano Scale SRAM lifetime", *IEEE 30th VLSI Test Symposium.*, 2012.

[13] D. Rodopoulos, et al, "Time and Workload Dependent Device Variability in Circuit Simulations" *Proc. Intl. Conf on IC Design and Technology*, pp: 1-4, 2011.

[14] M. T. Luque, et al, "From Mean Values to Distribution of BTI Lifetime of Deeply scaled FETs through Atomistic Understanding of the Degradation" *Sym. on VLSI Technology*, pp: 152-153, 2011.

[15] T. Sakurai et al, "Alpha-Power law MOSFET model and its applications to CMOS delay and other formulas", *IEEE JSSC*, Vol.25, No.2, April 1990,

[16] A. T. Krishnan, et al, "NBTI impact on transistor and circuit: Models, mechanisms and scaling effects", *IEDM*, 2003.

[17] B. Cheng, A. R. Brown, "Impact of NBTI/PBTI on SRAM Stability Degradation", *IEEE ELECTRON DEVICES LETTERS*, 2011.

[18] S. Kumar et al, "Impact of NBTI on SRAM Read Stability and Design for Reliability", *ISQED*, pp: 212-128, 2006.

[19] A. Carlson, "Mechanism of Increase in SRAM VMIN Due to Negative-Bias Temperature Instability", *IEEE TDMR*, 2007.

[20] A. Bansal et al, "Impact of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability", *JMR*, 2009.

[21] S. Khan, et al, "Impact of Partial Resistive Defects and Bias Temperature Defects and Bias Temperature Instability on SRAM Decoder Reliability", *Pro. of 7th IDT*, 2012.

[22] D. Rodopoulos, et al, "Atomistic Pseudo-Transient BTI Simulation with Inherent Workload Memory", *IEEE TDMR*, June 2014.

[23] S. Khan, et al, "Bias Temperature Instability Analysis of FinFET based SRAM cells", *DATE*, 2014.

[24] R. Menchaca et al, "Impact of Transistor Aging Effects on Sense Amplifier Reliability in Nano-Scale CMOS", *13th ISQED*, 2012.

[25] I. Agbo et al, "BTI Impact on SRAM Sense Amplifier", *8th IDT*, 16-18 Dec. 2013.

[26] H. Kukner et al, "Comparison of Reaction-Diffusion and Atomistic Trap-based Models for Logic Gates", *IEEE TDMR*, 2013.

[27] V. Chandra, R. Aitken, "Impact of Voltage Scaling on Nanoscale SRAM Reliability", *DATE*, 2009.

[28] S. Hamdioui, "Testing Static Random Access Memories: Defects, Fault Models and Test Patterns", *Kluwer Academic Press Publishers, The Netherlands*, 2004.

[29] H. Kukner et al, "BTI reliability from Planar to FinFET nodes: Will the next node be more or less reliable", *MEDIAN*, 2014.

[30] S. Cosemans, "Variability-aware design of low power SRAM memories", *Ph.D Thesis Katholieke Universiteit Leuven*, 2009.

[31] Predictive Technology Model "http://ptm.asu.edu/".

1444 IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL. 25, NO. 4, APRIL 2017

Integral Impact of BTI, PVT Variation, and Workload on SRAM Sense Amplifier

Innocent Agbo, *Student Member, IEEE*, Mottaqiallah Taouil, *Student Member, IEEE*, Daniël Kraak, *Student Member, IEEE*, Said Hamdioui, *Senior Member, IEEE*, Halil Kükner, *Member, IEEE*, Pieter Weckx, *Member, IEEE*, Praveen Raghavan, *Member, IEEE*, and Francky Catthoor, *Fellow, IEEE*

***Abstract*— The CMOS technology scaling faced over the past recent decades severe variability and reliability challenges. One of the major reliability challenges is bias temperature instability (BTI). This paper analyzes the impact of BTI on the sensing delay of standard latch-type sense amplifier (SA), which is one of the critical components of high performance memories; the analysis is done by incorporating the impact of process, voltage, and temperature variations (in order to investigate the severity of the integral impact) and by considering different workloads and four technology nodes (i.e., 45, 32, 22, and 16 nm). The results show the importance of taking the SA degradation into consideration for robust memory design; the SA degradation depends on the application and technology node, and the sensing delay can increase with 184.58% for the worst case conditions at 16 nm. The results also show that the BTI impact for nominal conditions at 16 nm reaches a 12.10% delay increment. On top of that, when extrinsic conditions are considered, the degradation can reach up to 168.45% at 398 K for 16 nm.**

***Index Terms*— Bias temperature instability (BTI), negative BTI (NBTI), positive BTI (PBTI), process variations, static RAM (SRAM) sense amplifier (SA).**

I. INTRODUCTION

TECHNOLOGY scaling poses major reliability challenges due to both intrinsic and extrinsic variations.

The variability of intrinsic device parameters are worsening in each CMOS technology generation; this is the result of unavoidable imperfections during fabrication and the introduction of new materials [1]. In addition, extrinsic variations in V_{dd} and temperature are consequences of changing operational and environmental conditions and also impact the transistors. On top of them, the intrinsic degradation of the devices causes major reliability challenges [2]–[4]. Bias temperature instability (BTI) [i.e., negative BTI (NBTI) in pMOS transistors and positive BTI (PBTI) in nMOS transistors] is a major reliability failure mechanism that affects the performance of MOS transistors by increasing their threshold voltage and reducing their drain current (I_d) over the operational lifetime [5], [9]. Hence, it impacts the robustness of both logic (such as CPUs) and memory [such as static RAM (SRAM)] circuits. Hence, analyzing the integral impact of all of these variables is needed in order to quantify the impact and degradation appropriate for the design for reliability solutions.

SRAM occupies a large fraction of semiconductor chip and plays a major role in the silicon area, performance, and critical robustness [10]. An SRAM system consists of an array of cells, its peripheral circuits such as row and column address decoders, control circuits, write drivers, and sense amplifiers (SAs).

Designing an optimal reliable memory system requires a deep understanding of the way it degrades in order to provide appropriate design-for-reliability schemes.

Many publications analyzed the BTI impact on SRAM cell array, while very limited work is published on the SRAM peripheral circuitry. Cheng and Brown [11] investigated the NBTI impact on static noise margin (SNM) and write noise margin degradation of the 6T SRAM cell. Kumar *et al.* [12] analyzed the impact of NBTI on the read stability and SNM of SRAM cells. Carlson [13] investigated the mechanism of NBTI degradation on SRAM metrics such as SNM. Kang *et al.* [14] studied the estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance. Weckx *et al.* [15] analyzed the implications of BTI-induced time-dependent statistics on yield estimation of digital circuits. Bansal *et al.* [16] presented insights into the stability of an SRAM cell under the worst case conditions and analyzed the effect of NBTI and PBTI, individually and collectively. Rodopoulos *et al.* [17] investigated the atomistic pseudotransient BTI simulation with built-in workloads. Wang *et al.* [18] investigated the statistical reliability analysis of NBTI impact on FinFET SRAMs and mitigation technique using independent gate devices. Khan *et al.* [19] investigated the BTI analysis of FinFET-based SRAM cell.

On the other hand, few authors have focused on reliability analysis of the SRAM peripheral circuit. Khan *et al.* [20] investigated the impact of partial opens and BTI on an SRAM address decoder.

In addition to that, Menchaca and Mahmoodi [21] analyzed the BTI impact on different SA designs implemented on a 32-nm technology node using failure probability (i.e., flipping a wrong value) as a reliability metric. Agbo *et al.* [22]–[25] investigated the BTI impact on the SRAM drain-input latch-type SA design implemented on 90, 65, and 45 nm for different

Manuscript received July 27, 2016; revised October 24, 2016; accepted November 30, 2016. Date of publication January 16, 2017; date of current version March 20, 2017.

I. Agbo, M. Taouil, D. Kraak, and S. Hamdioui are with the Computer Engineering Laboratory, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: i.o.agbo@tudelft.nl; m.taouil@tudelft.nl; d.h.p.kraak@tudelft.nl; s.hamdioui@tudelft.nl).

H. Kükner is with GLOBALFOUNDRIES, Singapore 738406 (e-mail: selahaddinhalil.kukner@globalfoundries.com).

P. Weckx, P. Raghavan, and F. Catthoor are with IMEC vzw, 3001 Leuven, Belgium, and also with Katholieke Universiteit Leuven, 3001 ESAT, Belgium (e-mail: pieter.weckx@imec.be; ragha@imec.be; francky.catthoor@imec.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2016.2643618

supply voltages using sensing delay and sensing voltage as reliability metrics. Agbo *et al.* [26] investigated a comparative BTI impact for SRAM cell and SA designs, while considering different applications and using swing delay and sensing delay as evaluation metrics.

The above statement clearly shows that most of the works focused on the SRAM cell. In addition, incorporating both variability and reliability are not explored very well. It is worth noting that understanding and quantifying the aging rate of each memory part/component is needed for an optimal reliable memory design; this is because the different parts may degrade with different rates depending on the workload (application).

This paper analyzes the degradation of the SA, a critical component of a memory especially for high-performance applications. This paper focuses on the standard latch-type SA design due to its superior performance [27] and analyzes process, voltage, and temperature (PVT) variations in combination with BTI for different workloads and technology nodes. The main contributions of this paper are as follows:

1) investigation of the BTI impact on the SA's sensing delay using eight realistic workloads;
2) thorough quantitative analysis of the BTI impact on the SA for four technology nodes, i.e., 45, 32, 22, and 16 nm;
3) investigation of the BTI impact on different workloads for different technology nodes;
4) different supply voltage impact analyses on SRAM SA sensing delay for different workloads;
5) investigation of different temperature impacts on SA for various technology nodes;
6) analysis of the integral impact of BTI and process variation on SA's sensing delay.

The rest of this paper is organized as follows. Section II introduces the functional model of SRAM system, the standard latch-type SA, the BTI mechanism and its model, and the variations targeted in this paper. Section III provides our analysis framework and analysis metric, and it also presents the performed experiments. Section IV reports, analyzes, and discusses the results. Finally, Section V concludes this paper.

II. BACKGROUND

This section presents the electrical model of the SA considered in this paper.

Subsequently, it presents the BTI mechanism and its model. Finally, it models the process and environmental variations considered in this paper.

A. SRAM Sense Amplifier

Several implementations of SAs have been proposed. In this paper, the standard latch-type SRAM strobed SA will be addressed, which is representative of industrial SA designs [27].

The structure of such an SA is depicted in Fig. 1. The width/length ratio of each transistor is presented by W/L. The operation of the SA consists of two phases. In the first phase, when SAenable is low, the access transistors *Mpass* and *MpassBar* connect to the *BL*(*BLBar*) with the internal

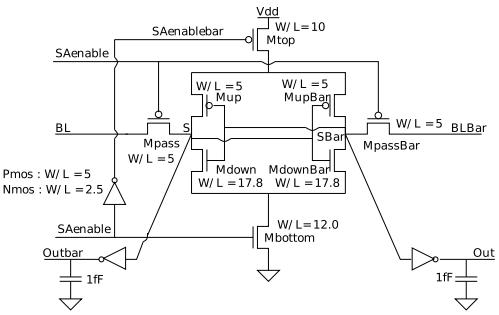


Fig. 1. Standard latch-type SA.

nodes *S*(*SBar*). In this phase, *Mtop* and *Mbottom* transistors are switched OFF. In the second phase, when SAenable is high, the pass transistors (i.e., *Mpass* and *MpassBar*) disconnect the *BL* and *BLBar* inputs from the internal nodes. The cross-coupled inverters get their current from Mtop and Mbottom and subsequently amplify the difference between *S* and *SBar* and produce digital outputs on *out* and *outbar*. The *S*(*SBar*) node is actively pulled down when *SBar*(*S*) exceeds the threshold voltage of Mdown. The positive feedback loop ensures low amplification time and produces the read value at its output. Moreover, all current paths are disabled when *S*(*SBar*) is at 0 V and *SBar*(*S*) is at V_{dd} or vice versa. This process is repeated for each read operation.

B. Bias Temperature Instability

The BTI mechanism takes place inside MOS transistors and causes a threshold voltage shift that negatively impacts the delay; its mechanism is described in the following section.

1) BTI Mechanism: BTI increases the absolute V_{th} value in MOS transistors. For pMOS, NBTI reduces the V_{th}, while for nMOS, PBTI increases the V_{th}.

Recently, exhaustive efforts have been put to understand NBTI [5], [9], [29]. Alam and Mahapatra [5] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion (RD) process. Nevertheless, this model seems to fail to explain the BTI degradation in the recovery phase [28]. Kaczer *et al.* [29] have analyzed NBTI using an atomistic model. The difference between the RD model and atomistic model is that the latter is based on the calibrations and its stochastic nature, while RD is based on deterministic nature. In [30], both the atomistic and RD models have been studied and compared. The results revealed that RD and atomistic models show similar trends. When only long-term reliability effects (across years) need to be studied, then the RD model is more suitable than the atomistic model because of its lightweight nature. In the context of this paper, this is our focus, so we opted for the RD model published in [5]. The model has two BTI phases, i.e., the stress phase and the relaxation phase.

Stress phase: In the stress phase, the silicon hydrogen bonds (≡Si–H) break at the silicon–oxide interface. The broken silicon bonds (≡Si−) remain at the interface (known as

interface traps) and the released H atoms/molecules diffuse toward the gate oxide. The number of interface traps (N_{IT}) generated after applying a stress of time (t) is given by [5]

$$N_{IT}(t) = \left(\frac{N_o \cdot k_f \cdot E_{ox}}{k_r}\right)^{2/3} \cdot \left(\frac{k_H}{k_{H_2}}\right)^{1/3} \cdot (6 \cdot D_0 \cdot t)^{1/6} \quad (1)$$

where N_o, k_f, k_r, k_H, and k_{H_2} represent the initial ≡Si–H density, ≡Si–H breaking rate, ≡Si− recovery rate, H to H_2 conversion rate, and H_2 to H conversion rate inside the oxide layer, respectively, while $E_{ox} = (V_{gs} - V_{th})/T_{ox}$ is the electric field across the gate oxide, which causes the breaking of silicon hydrogen bond at the interface, which is associated with k_f [6]–[8]. D_0=$D_{H_2} \cdot \exp(-E_A/kT)$ [31] is the diffusion coefficient of the produced H_2 species and E_A is the activation energy, k is the Boltzmann constant, and T is the temperature in Kelvin.

Relaxation phase: In the relaxation phase, there is no ≡Si–H breaking. However, the H atoms/molecules diffuse back toward the interface and anneal the ≡Si− bonds. The number of interface traps that *do not* anneal by the approaching H atoms during the relaxation phase is given by [12]

$$N_{IT}(t_o + t_r) = \frac{N_{IT}(t_o)}{1 + \sqrt{\frac{\xi \cdot t_r}{t_o + t_r}}} \quad (2)$$

where $N_{IT}(t_o)$ is the number of interface traps at the start of the relaxation, ξ is a relaxation coefficient with $\xi = 0.5$ [32], t_o is the duration of the previous stress phase, and t_r is the relaxation duration.

Threshold voltage increment: N_{IT} opposes the gate voltage, which results in a threshold voltage increment (ΔV_{th}). The relation between N_{IT} and ΔV_{th} is given by [33]

$$\Delta V_{th} = (1 + m) \cdot q \cdot N_{IT} / C_{ox} \cdot \chi \quad (3)$$

where m, q, and C_{ox} are the holes/mobility degradation that contributes to the V_{th} increment [34], electron charge, and oxide capacitance, respectively. χ is a BTI coefficient with a value $\chi = 1$ for NBTI and $\chi = 0.5$ for PBTI [32]. CMOS technology scaling results in an apparent oxide field increment and this speeds up the covalent bond breaking phenomenon of BTI impact. Therefore, it is essential to evaluate the reliability of these technologies.

Delay increment: BTI-induced ΔV_{th} of each individual MOS transistor has its contribution to the additional delay. A generalized first-order equation that relates BTI-induced ΔV_{th} in a transistor to dataline/output signal delay is given by [33], [35]

$$\Delta D = \frac{n \cdot \Delta V_{th}}{(V_{gs} - V_{th})} \quad (4)$$

where n is the velocity saturation index of majority carriers in MOS channels. Since NBTI causes ΔV_{th} to the pMOS transistor and PBTI causes ΔV_{th} to the nMOS transistor, this paper considers the threshold voltage shifts for both types of MOS transistors.

C. Variations

Variability has caused significant drifts from the predicted specification for a chip. These can be classified into intrinsic process, extrinsic environmental, and aging variations. Each is described next.

D. Process Variations

Process variation is an important concern for the SRAM SA's sensing delay [36] and caused by the imperfect circuit manufacturing process. They affect several transistor parameters including the effective channel length (L_{eff}), oxide thickness (t_{ox}), dopant concentration (N_a), and transistor width (w). There are two sources of variation: systematic variation and random variation. In this paper, only random variations are considered. The random variation can be described by a probability distribution. Here, we focus only on the V_{th} variation at time zero, which is the most important [36]. The standard deviation (SD) of the V_{th} shift is given by

$$\sigma_{V_{th0}} = \frac{A_{\Delta V_{TH}}}{\sqrt{2WL}} \quad (5)$$

where $A_{\Delta V_{TH}}$ is Pelgrom's constant [37], and W and L are the transistor width and length, respectively. The integration of two independent random V_{th} values result in factor 2 at the denominator.

E. Environmental Variations

Variations also occur due to environmental sources. We focus on the impact of temperature and supply voltage in this paper.

1) Temperature Variation: The temperature parameter fluctuations impact the operating condition of MOS transistors; the dependence of threshold voltage and temperature (VT) is given by [38]

$$V_{th} = C_{vt} - \frac{Q_{ss}}{C_o} + \Phi_{ms} \quad (6)$$

where C_{vt} is a constant that represents Fermi potential, the surface charge is Q_{ss} at the Si–SiO_2 interface, the gate oxide capacitance is C_o, and work function difference Φ_{ms} is a function of the temperature [38]

$$\Phi_{ms} = -0.61 - \Phi_F(T). \quad (7)$$

Here, $\Phi_F(T)$ is Fermi potential. Expression (8) shows that the work function difference reduces with respect to an increase in temperature and therefore leads to a threshold voltage decrement.

2) Supply Voltage Variation: The supply voltage fluctuations affect the operating speed of MOS transistors [36]. For example, variations in switching activity lead to uneven power/current demand across the die/circuitry and may cause logic failures [36]. In addition, transistor subthreshold leakage variations create an uneven load on the supply voltage network. A reduction in the supply voltage degrades the performance of the circuit/transistors and an increment in supply voltage enhances the performance.

F. Aging Variations

As already mentioned, BTI is one of the major mechanisms causing the aging of chips and induces V_{th} shifts. However, the BTI-induced V_{th} is a stochastic process for time $t > 0$.

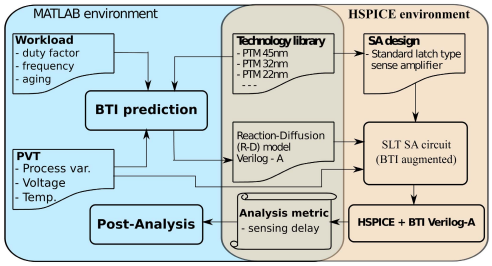


Fig. 2. Analysis framework for the standard latch SA circuit.

Si–H covalent bonds at the interface undergo stochastic fluctuations effected by random dopant fluctuations for time $t > 0$ (see equation 1).

The SD of the V_{th} variation caused by the interface traps is given by [14]

$$\sigma_{\Delta V_{\text{th}}(t)} = \sqrt{\frac{q t_{\text{ox}} \mu(\Delta V_{\text{th}})}{(1+m)\epsilon_{\text{ox}} A_G}} \tag{8}$$

where q is the charge, t_{ox} is the oxide thickness, μ is the mean of the BTI-induced threshold voltage shift, m is the mobility, ϵ_{ox} is the oxide electric field, and A_G is the effective area $(W * L)$.

Finally, the process and BTI-induced variation can be modeled together [18] using (5) and (8); the result is

$$\sigma_{V_{\text{th}}(t)} = \sqrt{\sigma^2_{V\text{th}0} + \sigma^2_{\Delta V_{\text{th}}(t)}}. \tag{9}$$

III. Analysis Framework

In this section, the analysis framework of the standard latch SA circuit is described. Thereafter, the output performance metric is presented. Finally, the conducted experiments are presented.

A. Framework Flow

Fig. 2 depicts a flexible and generic BTI framework for the standard latch-type SA circuit. The framework evaluates the BTI impact for different designs, technologies, and workloads under normal conditions and considering VT variations. The framework consists of MATLAB and HSPICE working environments. The MATLAB environment typically is used for preprocessing and postprocessing; it prepares BTI-augmented files to run in HSPICE. The results of HSPICE, which simulates the BTI-augmented netlist, are subsequently postprocessed in MATLAB. The analysis framework is divided into three blocks (i.e., input, processing, and output blocks) and they are explained next.

1) Input: The input blocks of the framework are the SA design, technology library, workload, and VT specification. They are explained as follows.

1) SA Design: In general, all SA designs can be used. In this paper, we focus only on the standard latch-type SA. The SA design is described by an HSPICE netlist. The differential input depends on the technology node. For example, implementing the above SA circuitry in 45 nm requires a differential input signal of 100 mV.
2) Technology Library: Different technology nodes are considered in this paper; they are 45, 32, 22, and 16 nm and are obtained from PTM library cards [39].
3) PVT: This block defines the process, temperature, and voltage variations. In this paper, we restrict ourselves to the random or local variations of the threshold voltage and use 1000 Monte Carlo runs for each experiment. We consider temperatures $T_1 = 298$ K, $T_2 = 348$ K, and $T_3 = 398$ K and supply voltages $V_1 = -10\%V_{\text{dd}}$, $V_2 = V_{\text{dd}}$, and $V_3 = +10\%V_{\text{dd}}$. Note that each technology has its own nominal voltage.
4) Workload: The shift in threshold voltage is a function of stress and relaxation durations of the transistors. This implies that BTI degradation depends on the number of ON and OFF (idle) states of the input patterns, which translates into workload. To perform realistic workload analysis, we assume that today's application consists of 10%–90% memory instructions and the percentage of read instructions is typically 50%–90%. We derive from this the following workload sequences: $S1$: $R0I^{99}$, $S2$: $R0R1I^{198}$, $S3$: $R0^4I^1$, $S4$: $(R0R1)^4I^2$, $S5$: $(R0)I^{24}$, $S6$: $(R0R1)I^{24}$, $S7$: $(R0)I^{50}$, and $S8$: $(R0R1)I^{50}$. In these sequences, $R0$ stands for read 0, $R1$ stands for read 1, and I for idle operation (which includes memory write operations). For example, $S1$: $R0I^{99}$ is the workload where read 0 is followed by 99 idle operations. The best and worst case sequences (i.e., $S2$ and $S3$, respectively) will be analyzed in most detail.

The workload inputs are typically characterized by their duty factor, frequency, and aging (or stress time). The BTI impact sensitivity is highly dependent on the input stimulus clock cycle (i.e., frequency), its aging, and duty factor (dc stress or ac stress) with respect to the affected device or circuitry.

1) *Frequency:* The BTI-induced degradation depends on the signal frequency to the SA design. In this experiment, the frequencies considered for the SA design are 1.32, 1.89, 2.70, and 3.86 GHz for the 45-, 32-, 22-, and 16-nm technology nodes, respectively. A design version is generated for each targeted frequency by scaling the device technology with a scale factor >= 0.7 and by selecting appropriate device dimensions [40].
2) *Aging:* The BTI-induced degradation strongly depends on the stress time. The stress time defines how long the workload sequence is being applied. A workload sequence is assumed to be repeated until the age time is reached.
3) *Duty Factor:* The input signal is a function of the duty factor that affects the BTI-induced degradation of the SA design. The duty factor of the input SAenable signal (see Fig. 2.) is approximately 0.48. This applies only to read operations. During write or idle operations, the SAenable signal is disabled. From the waveform analysis, we extract for all transistors individually their stress and relaxation cycles, thereby obtaining accurate duty

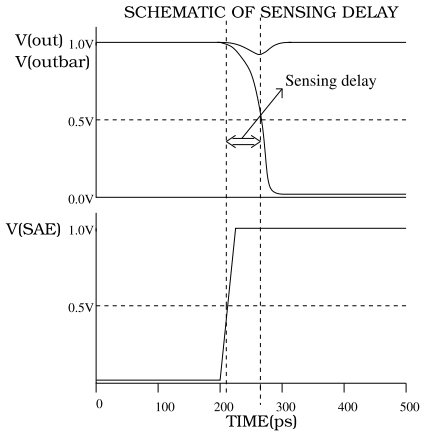


Fig. 3. Metric diagram of sensing delay.

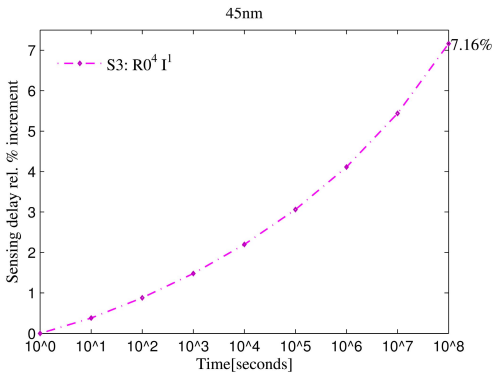


Fig. 4. BTI impact on Sensing delay for the 45-nm technology.

cycles for each transistor. This enhances the accuracy of our simulation results. Based on the duty factor and age time, interface traps ((3), (2)) or threshold voltage increments (3) can be attributed to all transistors in an accurate manner.

2) Processing: There are two processing blocks, the BTI predictor and the HSPICE simulation unit. The long-term BTI predictor uses the duty factor, frequency, and aging to predict the interface traps/threshold voltage increment of each device in the SA. In addition to workload inputs, inputs are required from the RD model (such as K_f, K_r, and D_H), technology parameters, and the VT. Once the BTI-induced V_{TH} increments are calculated per transistor, the original BTI-free netlist will be updated. This new netlist is simulated in HSPICE/Verilog-A.

3) Output: Finally, postanalysis of the results is performed for varying voltages and temperatures using the MATLAB environment.

B. Output Analysis Metrics

In this section, the sensing delay metric used for analyzing the BTI impact on SA is described.

1) Sensing Delay: The sensing delay metric is determined when the trigger signal (i.e., SAenable input signal) reaches 50% of the supply voltage and the target (i.e., either *out* or *outbar* falling output signal) reaches 50% of the supply voltage. The difference between the target and the trigger results in sensing delay as shown in Fig. 3. Furthermore, the relative variation of the sensing delay due to BTI is the difference between the measured sensing delay when BTI is added and referenced sensing delay when BTI is not added.

C. Simulation-Based Experiments Performed

In this paper, six sets of experiments are performed to analyze BTI impacts. These experiments are described as follows.

1) *Temporal BTI Impact Experiments:* The BTI impact on sensing delay of the SRAM SA is investigated.
2) *Workload-Dependent BTI Impact Experiments:* The BTI impact on the sensing delay of the SRAM SA for different workloads on different technology nodes is investigated.
3) *Technology-Dependent BTI Impact Experiments:* The BTI impact on sensing delay of the SRAM SA synthesized from different technology nodes is investigated.
4) *Supply-Voltage-Dependent Experiment:* The BTI impact on sensing delay of the SRAM SA for three supply voltages (i.e., from -10% of V_{dd} to V_{dd} and $+10\%$ of V_{dd}) for different technology nodes is presented.
5) *Temperature-Dependent Experiments:* The BTI impact on sensing delay of the SRAM SA for three temperatures (i.e., 298 K, 348 K, and 398 K) for different technology nodes is explored.
6) *Combined Impact of BTI- and PVT-Dependent Experiments:* The combined BTI and PVT impact on the SA sensing delay for the best and worst case workloads for different technology nodes, supply voltages, and temperatures is analyzed.

IV. SIMULATION RESULTS

In this section, we present the analysis results of the experiments mentioned in the previous section.

A. Temporal BTI Impact

Fig. 4 shows the BTI impact of workload *S*3 on the sensing delay for different aging times for nominal supply VT. Fig. 4 shows that the sensing delay increases as the SA ages. For this workload, the sensing delay increases from 14.10 to 15.11 ps (7.16% increase) caused by the BTI only.

B. Workload-Dependent BTI Impact

The BTI-induced degradation is sensitive to the applied workloads.

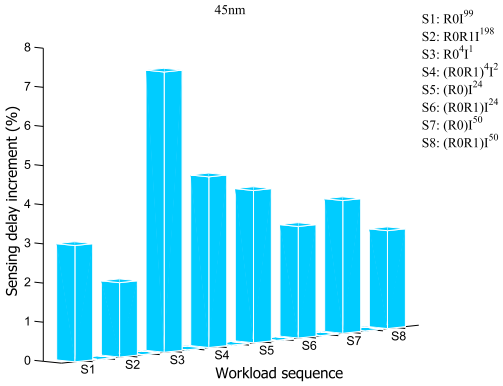


Fig. 5. Workload-dependent sensing delay.

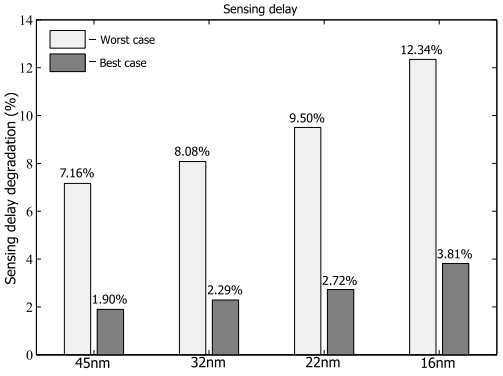


Fig. 6. BTI impact on sensing delay for all technology nodes.

Fig. 5 shows the BTI-induced degradation of the SA for the eight workloads defined in Section III at nominal supply VT for 3 operational years.

Fig. 5 shows a large variation in the relative sensing delay increment. For instance, workload $S2$ is impacted the lowest by BTI (1.90% degradation), whereas $S3$ the most (7.16% degradation). This can be explained by the severity of the workload. We define $S2$ and $S3$ as the best and worst case workloads, respectively.

C. Technology-Dependent BTI Impact

Fig. 6 depicts the relative BTI impact on the sensing delay for different technology nodes for the best case and worst case workloads at nominal supply VT. Note that the nominal supply voltage equals $V_{dd} = 1.0$ V for 45 nm, $V_{dd} = 0.9$ V for 32 nm, $V_{dd} = 0.8$ V for 22 nm, and $V_{dd} = 0.7$ V for 16 nm.

Fig. 6 shows that the sensing delay degradation is obviously dependent on the applied workload. For example, for the 22-nm technology node, the BTI-induced degradation is 9.50% for the worst case workload and only 2.72% for the best case workload, respectively.

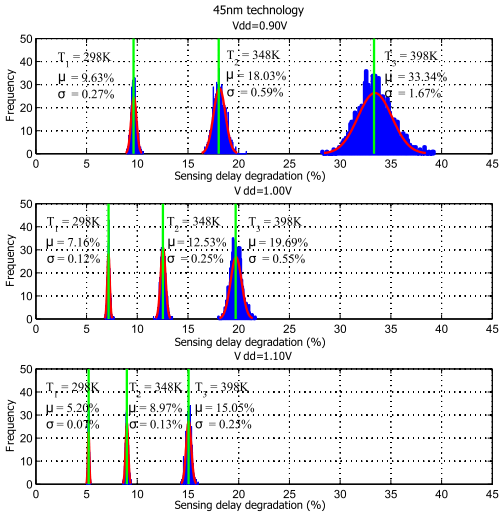


Fig. 7. Worst case sensing delay for supply VT variations for the 45-nm technology.

Fig. 6 also shows that the sensing delay degradation increases for smaller feature sizes irrespective of the two workloads. For instance, for the worst case workload, the degradation is 7.16% for the 45-nm technology node, 8.08% for the 32-nm technology node, 9.50% for the 22-nm technology node, and 12.34% for the 16-nm technology node. Especially at the lower nodes, the BTI may significantly impact the device reliability.

D. Supply-Voltage-Dependent BTI Impact

Figs. 7 and 8 depict the BTI-induced sensing delay for the 45-nm technology node, for the worst case and best case workloads at various supply voltages and temperatures. Figs. 7 and 8 show for each VT experiment the impact on the sensing delay (on the horizontal axis). Note that only 1000 Monte Carlo simulations have been performed for each experiment due to computational constraints. The 1000 simulations are grouped in 50 bins (see the blue curves in Figs. 7 and 8) with their frequency being plotted on the vertical axis. On top of these bins, a normal distribution is fit (see the red curves in Figs. 7 and 8), which reflects and follows the ΔV_{th} distribution due to BTI degradation. The vertical green lines in Figs. 7 and 8 present the mean degradation of the sensing delay (in percentage for each experiment). For example, for the worst case workload with $V_{dd} = 0.9$V and temperature $T_1 = 298$ K, the mean sensing delay $\mu = 9.63\%$ and the SD $\sigma = 0.27\%$. In this section, we discuss only the results of the voltage experiments for nominal temperature $T_1 = 298$ K.

From Figs. 7 and 8, we conclude that the BTI-induced degradation is significantly impacted by the voltage variations and workload. This is true for both the mean and SD. For instance, for the worst case workload (at T_1), the BTI-induced

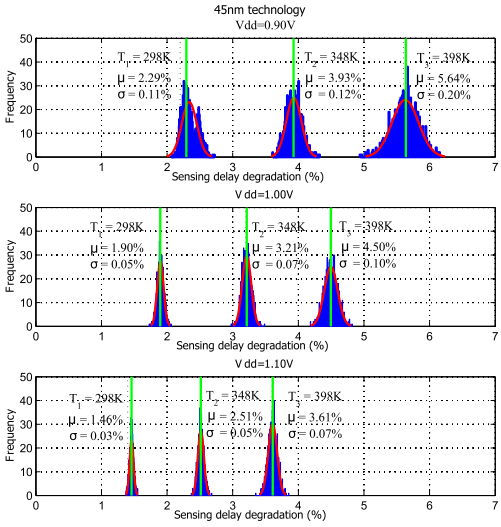


Fig. 8. Best case sensing delay for supply VT variations for the 45-nm technology.

mean degradation on sensing delay is 9.63% at 0.9 V, while only 5.20% at 1.1 V. For the best case workload, these values are 2.29%, 1.90%, and 1.46%, respectively. Moreover, the SD of the sensing delay for the worst case workload at T_1 is 0.27% at 0.9 V, while 0.12% at 1.0 V and 0.07% for $V_{dd} = 1.1$ V. For the best case workload, the SDs of the degradation equally are 0.11%, 0.05%, and 0.03% for supply voltages 0.9 V, 1.0 V, and 1.1 V, respectively. Hence, at a higher voltage, the *relative* degradation and its spread are lower for both workloads. Note that it is about relative impact and not the absolute one. The absolute impact at higher voltage is larger than at lower voltage.

Fig. 8 shows similar data as Fig. 7 but for the best case workload. We observe similar trends as for the worst case workload. The mean and SD of the degradation reduces for higher V_{dd} (for temperature fixed at $T_1 = 298$ K). However, they differ in amplitude with respect to the best case workload. For example, the mean and SD at T_1 and $V_{dd} = 0.90$V are 2.29% and 0.11%, respectively, for the best case workload, while 9.63% and 0.27%, respectively, for the worst case workload.

E. Temperature-Dependent BTI Impact

Figs. 7 and 8 also show the BTI impact for various temperatures. Note that the temperature is varied between 298 K and 398 K. Figs. 7 and 8 show that both the mean and SD of the degradation are significantly impacted by the temperature, especially at −10% supply voltage.

For example, for the worst case workload as shown in Fig. 7, the mean sensing delay degradation is 9.63% for $V_{dd} = 0.9$V and $T_1 = 298$ K and equals 18.03% for $T_2 = 348$ K and 33.34% for $T_3 = 398$ K. For the same conditions, the SD

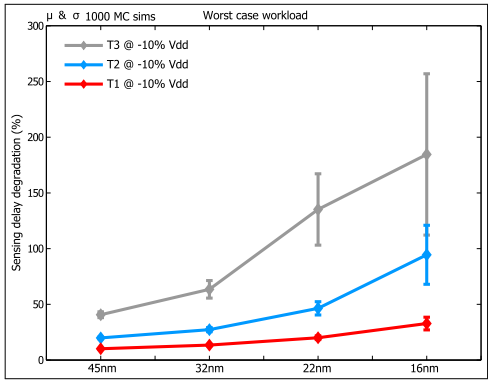


Fig. 9. Worst case SD for −10% Vdd for various nodes and temperatures.

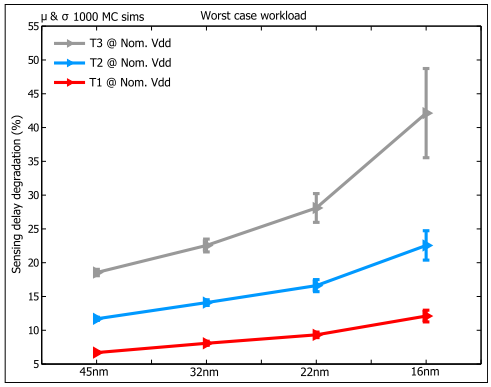


Fig. 10. Worst case SD for Nom. Vdd for various nodes and temperatures.

increases from 0.27% (at T_1) to 1.67% (at T_3). At a higher voltage, the same trends are observed with lower impact. For example, the mean degradations at $V_{dd} = 1.1$ V and T_1 equal 5.20% and 8.97% at T_3.

From this, we deduce that the lower the operational voltage, the higher the impact of temperature. Similar observations can be made from Fig. 8 for the best case workload.

F. Combined Impact of BTI- and PVT-Dependent Experiments

Figs. 9–11 depict the sensing delay for the worst case workload for the three supply voltages, respectively. In Figs. 9–11, our analysis focused on both the mean μ (denoted by the opaque markers) BTI-induced impact and the SD σ (denoted by the edges of the vertical lines). Figs. 9–11 show the impact of technology and temperature for each voltage value. Note that Figs. 7 and 8 depict only the VT impact for the 45-nm technology.

From Figs. 9–11, we conclude the following.

1) For newer technology, the BTI impact worsens; this conclusion is in line with Fig. 6. This conclusion can be

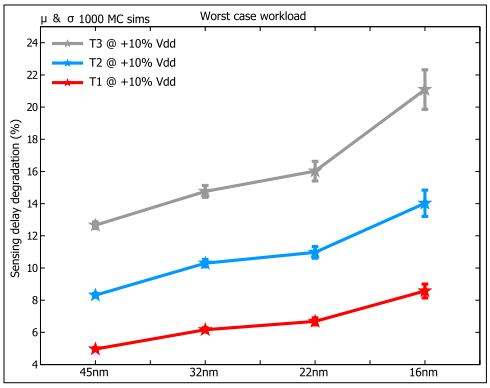


Fig. 11. Worst case SD for +10% Vdd for various nodes and temperatures.

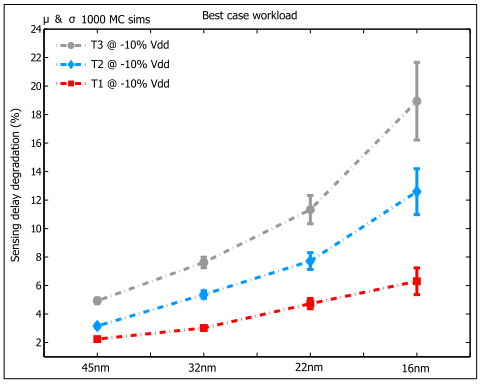


Fig. 12. Best case SD for −10% Vdd for various nodes and temperatures.

TABLE I
FAILING POINTS AT −10% V_{dd} FOR WORST CASE WORKLOAD

		Temperature		
		T1	T2	T3
Technology	$22nm$	0	0	663
	$16nm$	0	442	986

easily made as all graphs are monotonically increases. However, the relative degradation increases with a higher temperature. For example, after an operation of 10^8 s for temperature $T1 = 298$ K, the BTI-induced mean degradation is 10.13% for 45 nm, while 13.42%, 20.02%, and 32.78% for 32, 22, and 16 nm, respectively. In addition, for temperature $T3 = 398$ K, the degradation is up to 4.00× more for 45 nm, while 4.73×, 6.75×, and 5.60× more for 32, 22, and 16 nm, respectively. Furthermore, the mean sensing delay degradation reduces with 10% increase in supply voltage. For example, the degradation reduces for $T1 = 298$ K up to 1.51× for 45 nm, while 1.66× for 32 nm, and 2.15× for 22 nm, and only 2.71× for 16 nm. Moreover, for T3 = 398 K, the BTI impact degradation reduces up to 2.20× for the 45-nm node, 2.82× for the 32-nm node, 4.81× for the 22-nm node, and only 4.38× for the 16-nm node.

2) The SD is the highest for lower technology nodes, higher temperatures, and lower voltages. For example, after an operation of 10^8 s of $T1$, the degradation distribution is 0.27% for 45 nm, while 0.60%, 1.64%, and 5.69% for the 32-, 22-, and 16-nm nodes, respectively. On top of that, the SD for a higher temperature, i.e., $T3 = 398$ K, is up to 9.63× more for the 45-nm node, while 13.15×, 19.51×, and 12.72× for the 32-, 22-, and 16-nm nodes, respectively.

Table I provides information regarding failed simulation (i.e., signal flipped at the output). They only occurred for the advanced nodes (i.e., 22 and 16 nm) at −10% V_{dd} and at high temperatures (i.e., 348 K and 398 K). This shows that scaling and bad operating conditions may significantly impact the reliability. For example, 663 out of 1000 points failed at 22 nm for $T3$. At 16 nm, even failing points have been observed at $T2$, while 986 (which includes almost the entire set) points failed at $T3$.

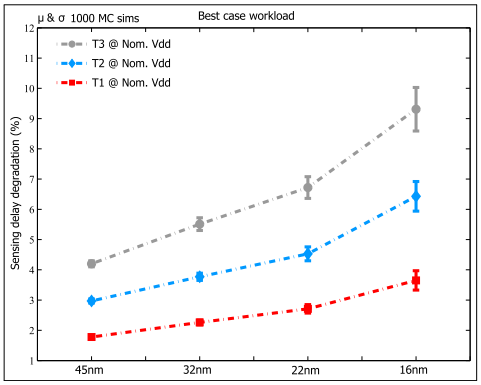


Fig. 13. Best case SD for Nom. Vdd for various nodes and temperatures.

Figs. 12–14 show similar trends but for the best case workload. The impact of this workload on the sensing delay is, however, much less. For example, the mean degradation reduces up to 4.52× at 298 K to 8.26× at 398 K for 45 nm, while this is only 4.44× to 8.33×, 4.23× to 11.93×, and 5.20× to 9.75× for the 32-, 22-, and 16-nm nodes, respectively. On top of that, the degradation deviation reduces up to 3.86× at 298 K to 13.00× at 398 K for 45 nm, while 3.75× to 21.32× for the 32-nm node, 4.56× to 32.32× for the 22-nm node, and only 6.05× to 26.52× for the 16-nm node. Furthermore, the mean sensing delay and deviation degradation reduces as the supply voltage rises by 10%. For example, the relative mean degradation also reduces up to 3.47× at 298 K to 3.56× at 398 K for 45 nm, while 3.06× to 3.04×, 3.23× to 3.03×, and 2.78× to 2.84× for

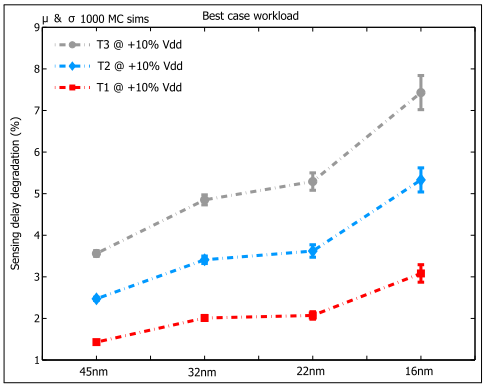


Fig. 14. Best case SD for +10% Vdd for various nodes and temperatures.

32, 22, and 16 nm, respectively. On top of that, the deviation degradation also reduces up to 2.33× at 298 K to 3.00× at 398 K, while this is only 2.17× to 3.08×, 2.44× to 2.90×, and 2.05× to 3.00× for the 32-, 22-, and 16-nm nodes, respectively.

G. Discussion

The SRAM SA robustness and reliability are critical for the overall design of memory systems. The present analysis shows that the BTI-induced sensing delay of standard latch-type SA is a function of process variations, varying supply voltages and temperatures, different technology nodes, and workloads. Evaluating the simulation results with respect to degradation, we observe the following.

1) The SA *sensing delay degradation* is clearly influenced or impacted by the application, *i.e., by the workloads* it is exposed to. In case an application is read intensive, a second parallel SA scheme can be exploited to minimize the performance loss, by sharing the read operations between the two SAs. This, however, doubles the area.
2) The BTI-induced degradation is more significant at lower nodes; this could lead to read failures at lower supply voltage. This suggests that there must be a tradeoff between performance and reliability.
3) A higher supply voltage reduces sensing delay's mean degradation irrespective of the workload and temperature. This indicates that a tradeoff is possible among BTI degradation, power consumption, and the latency.
4) A higher junction temperature increases sensing delay's mean degradation.
   Hence, proper cooling or reducing the SA activity may also reduce the degradation. For example, switching the SA OFF while using the secondary SA reduces the degradation impact.

Therefore, for future designs, SA designers need to consider proper safe margins for critical environmental conditions and long-term operations.

V. CONCLUSION

This paper investigated the combined impact of BTI, process variations, various supply voltages and temperatures, and different workloads on standard latch-type memory SA for different technologies. The results show that the sensing delay degradation is strongly workload dependent. Both the scaling and increase in temperature severely impact the BTI degradation. Increasing the supply voltage reduces the BTI-induced degradation leading to more reliable and robust SAs, but at the cost of a higher power consumption. Nevertheless, process and BTI variations reduce the reliability and may even cause failures. Optimizing a reliable memory system requires the consideration of all its subparts, i.e., their degradation rate depends on the application (e.g., workload and temperature), for instance, the aging rate of the SA may differ from those of the memory array and other peripheral circuits.

REFERENCES

[1] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variations and impact on circuits and microarchitecture," in *Proc. Design Autom. Conf. (DAC)*, 2003, pp. 338–342.
[2] ITRS. (2004). *International Technology Roadmap for Semiconductor 2004*. [Online]. Available: https://www.itrs.net/common/2004update/2004update.htm
[3] S. Borkar, "Microarchitecture and design challenges for gigascale integration," in *Proc. Int. Symp. Microarchitecture*, 2004, p. 3.
[4] S. Hamdioui, D. Gizopoulos, G. Guido, M. Nicolaidis, A. Grasset, and P. Bonnot, "Reliability challenges of real-time systems in forthcoming technology nodes," in *Proc. Design, Autom. Test Eur. Conf. Exhibit.*, 2013, pp. 129–134.
[5] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron. Rel.*, vol. 45, no. 1, pp. 71–81, 2005.
[6] A. E. Islam, H. Kufluoglu, D. Varghese, S. Mahapatra, and M. A. Alam, "Recent issues in negative-bias temperature instability: Initial degradation, field dependence of interface trap generation, hole trapping effects, and relaxation," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2143–2154, Sep. 2007.
[7] W. Wang, V. Reddy, A. T. Krishnan, R. Vattikonda, S. Krishnan, and Y. Cao, "Compact modeling and simulation of circuit reliability for 65-nm CMOS technology," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 4, pp. 509–517, Dec. 2007.
[8] W. Wang, S. Yang, S. Bhardwaj, S. Vrudhula, F. Liu, and Y. Cao, "The impact of NBTI effect on combinational circuit: Modeling, simulation, and analysis," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 2, pp. 173–183, Feb. 2010.
[9] S. Zafar *et al.*, "A comparative study of NBTI and PBTI (charge trapping) in SiO_2/HfO_2 stacks with FUSI, TiN, Re gates," in *Proc. VLSI Technol. Symp.*, Jun. 2006, pp. 23–25.
[10] P. Pouyan, E. Amat, and A. Rubio, "Process variability-aware proactive reconfiguration technique for mitigating aging effects in nano scale SRAM lifetime," in *Proc. IEEE 30th VLSI Test Symp.*, Apr. 2012, pp. 240–245.
[11] B. Cheng, A. R. Brown, and A. Asenov, "Impact of NBTI/PBTI on SRAM stability degradation," *IEEE Electron Devices Lett.*, vol. 32, no. 6, pp. 740–742, Jun. 2011.
[12] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," in *Proc. ISQED*, Mar. 2006, p. 218.
[13] A. Carlson, "Mechanism of increase in SRAM V_{MIN} due to negative-bias temperature instability," *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 3, pp. 473–478, Sep. 2007.
[14] K. Kang, S. P. Park, K. Roy, and M. A. Alam, "Estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCD)*, Nov. 2007, pp. 730–734.
[15] P. Weckx *et al.*, "Implications of BTI-induced time-dependent statistics on yield estimation of digital circuits," *IEEE Trans. Electron Devices*, vol. 61, no. 3, pp. 666–673, Mar. 2014.

[16] A. Bansal *et al.*, "Impacts of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability," *J. Microelectron. Rel.*, vol. 49, no. 6, pp. 642–649, 2009.

[17] D. Rodopoulos, P. Weckx, M. Noltsis, F. Catthoor, and D. Soudris, "Atomistic pseudo-transient BTI simulation with inherent workload memory," *IEEE Trans. Device Mater. Rel.*, vol. 14, no. 2, pp. 704–714, Jun. 2014.

[18] Y. Wang, S. D. Cotofana, and L. Fang, "Statistical reliability analysis of NBTI impact on FinFET SRAMs and mitigation technique using independent-gate devices," in *Proc. IEEE/ACM Int. Symp. Nanosc. Archit. (NANOARCH)*, Jul. 2012, pp. 4–6.

[19] S. Khan *et al.*, "Bias temperature instability analysis of FinFET based SRAM cells," in *Proc. Design, Autom. Test Eur. Conf. Exhibit. (DATE)*, Mar. 2014, pp. 24–28.

[20] S. Khan, M. Taouil, S. Hamdioui, H. Kukner, P. Raghavan, and F. Catthoor, "Impact of partial resistive defects and bias temperature instability on SRAM decoder reliablity," in *Proc. 8th IEEE Int. Design Test Symp.*, Dec. 2013, pp. 1–6.

[21] R. Menchaca and H. Mahmoodi, "Impact of transistor aging effects on sense amplifier reliability in nano-scale CMOS," in *Proc. 13th Int. Symp. Quality Electron. Design*, Mar. 2012, pp. 342–346.

[22] I. Agbo *et al.*, "Comparative BTI analysis for various sense amplifier designs," in *Proc. 19th Int. Symp. Design Diagnostics Electron. Circuits Syst. (DDECS)*, 2016, pp. 1–6.

[23] I. Agbo *et al.*, "Quantification of sense amplifier offset voltage degradation due to zero- and run-time variability," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, Jul. 2016, pp. 725–730.

[24] I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, and F. Catthoor, "BTI analysis of SRAM write driver," in *Proc. 10th Int. Design Test Symp. (IDT)*, 2015, pp. 100–105.

[25] I. Agbo, S. Khan, and S. Hamdioui, "BTI impact on SRAM sense amplifier," in *Proc. 8th IEEE Int. Design Test Symp.*, Dec. 2013, pp. 16–18.

[26] I. Agbo *et al.*, "Comparative BTI impact for SRAM cell and sense amplifier designs," in *Proc. MEDIAN Finale*, Nov. 2015, pp. 10–11.

[27] S. Cosemans, "Variability-aware design of low power SRAM memories," Ph.D. dissertation, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, 2009.

[28] V. Huard, "Two independent components modeling for negative bias temperature instability," in *Proc. Int. Phys. Rel. Symp. (IPRS)*, May 2010, pp. 33–42.

[29] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *Proc. Int. Phys. Rel. Symp. (IPRS)*, Apr. 2005, pp. 381–387.

[30] H. Kukner *et al.*, "Comparison of reaction-diffusion and atomistic trap-based BTI models for logic gates," *IEEE Trans. Device Mater. Rel.*, vol. 14, no. 1, pp. 182–193, Mar. 2014.

[31] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the dispersive versus arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications," in *IEDM Tech. Dig.*, Dec. 2005, pp. 1–4.

[32] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "An analytical model for negative bias temperature instability," in *Proc. ICCAD*, 2006, pp. 493–496.

[33] B. C. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *IEEE Electron Device Lett.*, vol. 26, no. 8, pp. 560–562, Aug. 2005.

[34] A. T. Krishnan, V. Reddy, S. Chakravarthi, J. Rodriguez, S. John, and S. Krishnan, "NBTI impact on transistor and circuit: Models, mechanisms and scaling effects," in *IEDM Tech. Dig.*, Dec. 2003, pp. 14.5.1–14.5.4.

[35] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[36] S. Sapatnekar, "Overcoming variations in nanometer-scale technologies," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 1, pp. 5–18, Mar. 2011.

[37] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[38] R. Wang, J. Dunkley, T. A. DeMassa, and L. F. Jelsma, "Threshold voltage variations with temperature in MOS transistors," *IEEE Trans. Electron Devices*, vol. 18, no. 6, pp. 386–388, Jun. 1971.

[39] *Predictive Technology Model*. accessed on Sep. 30, 2008. [Online]. Available: http://ptm.asu.edu

[40] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, Jul./Aug. 1999.

Innocent Agbo (S'10) received the M.Sc. degree in computer engineering from the Delft University of Technology, Delft, The Netherlands, where he is currently pursuing the Ph.D. degree in electrical engineering, mathematics, and computer science.

His current research interests include the reliability analysis, design, monitoring, and mitigation of memory systems.

Mr. Agbo received the Best Paper Award at the IEEE Computer Society Annual Symposium on VLSI 2016, and the Best Paper Award at the International conference on Design and Test of Integrated Systems in the nano-era DTIS 2015.

Mottaqiallah Taouil (S'10) received the M.Sc. and Ph.D. degrees (both with Hons.) in computer engineering from the Delft University of Technology (TU Delft), Delft, The Netherlands.

He is currently a Post-Doctoral Researcher with the Dependable Nano Computing Group, TU Delft. His current research interests include reconfigurable computing, embedded systems, VLSI design and test, built-in-self-test, 3-D stacked integrated circuits, architectures, design for testability, yield analysis, and memory test structures.

Daniël Kraak (S'16) was born in Dordrecht, The Netherlands, in 1992. He received the bachelor's degree in electrical engineering and the master's degree in computer engineering from the Delft University of Technology, Delft, The Netherlands, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree with the Computer Engineering Laboratory. His master's thesis was titled "Experimental and Industrial Evaluation of Variability Resilient Schemes" and was performed in cooperation with NXP Semiconductors, Eindhoven, the Netherlands.

His current research interests include robust memory design and mitigation for aging.

Said Hamdioui (M'99–SM'11) received the M.S.E.E. and Ph.D. degrees (both with Hons.) from the Delft University of Technology (TU Delft), Delft, The Netherlands.

He was with Intel Corporation, Santa Clara, CA, USA, Philips Semiconductors Research and Development, Crolles, France, and Philips/NXP Semiconductors, Nijmegen, The Netherlands. He is currently leading emerging dependable-nano computing research activities with the Computer Engineering Laboratory, TU Delft. He holds one patent, and has authored one book and co-authored over 170 conference and journal papers. His current research interests include testability, reliability, hardware security, and the emerging memristor-based computing.

Dr. Hamdioui is a member of the AENEAS/ENIAC Scientific Committee Council. He was a recipient of the European Design Automation Association Outstanding Dissertation Award 2001, the Best Paper Award at the IEEE Computer Society Annual Symposium on VLSI 2016, the 2015 HiPEAC Technology Transfer Award, the Best Paper Award at the 33rd IEEE International Conference on Computer Design ICCD 2015, the Best paper Award at the International conference on Design and Test of Integrated Systems in the nano-era DTIS 2011, and the IEEE Nano and Nano Korea Award at the IEEE NANO 2010. He currently serves on the editorial board of IEEE DESIGN & TEST, and Associate Editor of IEEE Transactions on VLSI Systems (TVLSI), and of the *Journal of Electronic Testing: Theory and Applications*.

Halil Kükner received the B.Sc. degree in electrical engineering from Sabanci University, Istanbul, Turkey, in 2008, and the M.Sc. degree in electrical engineering from Technische Universiteit Delft, Delft, The Netherlands, in 2010.

He is currently a Ph.D. Researcher with the Circuits and Systems at the ICT Group, Interuniversity Microelectronics Center, Leuven, Belgium, and also with the Department of Electrical Engineering, Katholieke Universiteit Leuven, Leuven. His current research interests include the reliability of digital circuits, microarchitecture, memory built-in self-testing, image and video processing, and motion estimation.

Mr. Kükner was a recipient of the Best Paper Award at DTIS.

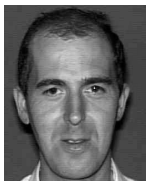
Pieter Weckx (M'14) received the M.Sc. degree in nanoscience and nanotechnology from Katholieke Universiteit Leuven, Leuven, Belgium, in 2011, where he is currently pursuing the Ph.D. degree in electronics and electrical engineering.

His current research interests include the modeling and simulation of time-dependent variability problems in nanoscaled electronic devices, statistical circuit simulations, and stochastic/deterministic clustering of circuit degradation behavior.

Praveen Raghavan (M'04) received the bachelor's degree in electrical engineering from the Regional Engineering College, Tiruchirappalli, India, the master's degree in electrical engineering from Arizona State University, Tempe, AZ, USA, and the Ph.D. degree from Katholieke Universiteit Leuven, Leuven, Belgium, in 2009.

In 2007, he joined the Berkeley Wireless Research Center, University of California, Berkeley, CA, USA, as a Visiting Researcher. He was the Lead Architect with imec, Leuven, Belgium, focused on multigigabit software-defined radio baseband chip set. He is currently a Principal Scientist with the Design Technology Enablement Group, imec, where he is the lead for investigating the impact of future technologies on circuits and systems. He has authored over 100 conference and journal papers, and holds more than 30 patents. His current research interests include processor architecture design, impact of deeply scaled technology on architectures/systems, reliability, variability, low-power design, and software defined radios.

Dr. Raghavan is currently with the Technical Committee of CODES + ISSS, DATE, and IEEE DISPS.



Francky Catthoor (F'05) received the engineering degree and the Ph.D. degree in electrical engineering from Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, in 1982 and 1987, respectively.

From 1983 to 1987, he was a Researcher, focused on VLSI design methodologies for digital signal processing, with Prof. H. De Man and Prof. J. Vandewalle as Ph.D. Thesis Advisors. Since 1987, he has been the Head of several research domains in the area of high-level and system synthesis techniques and architectural methodologies with the Systems Division, imec. Since 1989, he has been an Assistant Professor with the Department of Electrical Engineering, KU Leuven, where he has been a Full Professor (part-time) since 2000. He is currently a fellow with imec. His current research interests include architecture design methods and system-level exploration for power and memory footprint within real-time constraints, oriented toward data storage management, global data transfer optimization, and concurrence exploitation at Leuven, Belgium.

Dr. Catthoor was a member of the Steering Board for the VLSI Technical Committee of the IEEE Circuits and Systems Society in 1997. He received the Young Scientist Award from the Marconi International Fellowship Council in 1986. He was the Program Chair of the 1997 IEEE International Symposium on System Synthesis (ISSS) and the General Chair of the 1998 ISSS. He was also the Program Chair and Main Organizer of the 2001 IEEE Signal Processing Systems Conference. He was an Associate Editor of the IEEE TVLSI in 1995–1998, and the IEEE TRANSACTIONS ON MULTIMEDIA in 1999–2001. Since 2002, he has been an Associate Editor of the "*ACM Transactions on Design Automation for Embedded Systems*," and since 1996, he has been an Editor of Kluwer's the "*Journal of Very Large Scale Integration Signal Processing*." Since 1999, he has been on the Steering Board for the IEEE TVLSI.

2016 IEEE Computer Society Annual Symposium on VLSI

2

Quantification of Sense Amplifier Offset Voltage Degradation due to Zero- and Run-time Variability

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Pieter Weckx[1,2] Stefan Cosemans[1] Praveen Raghavan[1] Francky Catthoor[1,2] and Wim Dehaene[2]
[1]imec vzw., Kapeldreef 75, B-3001, Leuven, Belgium
[2]Katholieke Universiteit Leuven, ESAT, Belgium
{Pieter.Weckx, stefan.cosemans, Ragha, Francky.catthoor}@imec.be
wim.dehaene@esat.kuleuven.be

***Abstract*—Nowadays, typical (memory) designers add design margins to compensate for uncertainties; however, this may be overestimated leading to yield loss, or underestimated leading to reduced reliability designs. Accurate quantification of all uncertainties is therefore critical to provide high quality and optimal designs. These uncertainties are caused by zero-time variability (due to process variability), and by run-time variability (due to environmental variabilities such as voltage and temperature, or due to temporal variability such as aging). This paper uses an accurate methodology to predict the impact of both zero- and run-time variability on the offset voltage of sense amplifiers while considering different workloads and PVT variations for a pre-defined failure rate. The results show a marginal impact of environmental run-time variability on the offset specification when considering zero-time variability only, while this becomes significant (up to $2\times$) when incorporating aging run-time variability. The results can be used to quantify whether the required offset voltage is met or not for the targeted lifetime; hence, enable the designer to take appropriate measures for an efficient and optimized design, depending on the targeted application lifetime.**

***Index Terms*—Offset voltage, zero-time variability, run-time variability, SRAM sense amplifier**

I. INTRODUCTION

In recent decades, CMOS technology has been sustained with aggressive downscaling that poses major challenges on the reliability of devices [1–3]. The sources of this unreliability in todays technologies are mainly caused due to variability during manufacturing or at run-time [1]. As a result of the manufacturing process, devices will suffer from process variations, which changes the properties of the manufactured devices from the targeted ones. Hence, similar manufactured devices end up having different characteristics, referred to as *process* or *time-zero variation*. On the other hand, *run-time variations* cause the device properties to change and/or degrade during their operational lifetime. Such variations are mainly due to *environmental variations* such as supply voltage fluctuations and temperature variations, and *temporal* or *aging variations* such as *Bias Temperature Instability* [4]; both show a severe impact with CMOS scaling [1]. All these variations cause the devices to behave differently than intended, which may cause the devices (or circuits) to fail if appropriate measures are not taken. Designers usually use a conservative guardbanding and apply extra design margins [5] to ensure the correct operation for the worst-case variations during the targeted circuit lifetime. However, a pessimistic guardbanding leads to either yield or performance loss, while an optimistic guardbanding increases the test escapes and in-field failures. Therefore, an *accurate* estimation of the impact of *all kinds of* variations at circuit level (and also at the architecture level) is needed to obtain a high quality and optimal design. In this paper, we focus on the estimation of sense amplifier (SA) offset voltage using the integral impact of process and run-time variations, especially investigating the contribution of run-time variability due to aging to the overall impact as compared with zero-variability and environment variability; this guides the designers to optimize the guardbands and margins depending on the targeted product quality and application lifetime. Note that the SA delay, which is an integral part of the path delay and strongly related to the offset-voltage, plays a key role in defining memory design margins. For a fixed sensing delay, the higher the offset the more time needed for the read operation, as more time will be needed to discharge one of the cell's bitlines.

Only limited work has been presented for the characterization of the offset voltage in SAs. In [6], the authors presented a tunable SA to cope with with-in die variations; the authors estimated the offset voltage at design time based on process variations. In [7], the authors characterize the SA Input Offset by a physical circuit monitoring (implemented in real silicon) in order to estimate the yield. In [8], the authors presented a scheme to determine the signal margins for DRAM SAs based on offset distribution measurements. Prior work mainly focused on time-zero variation. Methods to estimate the SA offset voltage in the presence of both manufacturing and run-time variability at design stage are still missing.

This paper uses an *accurate* method to estimate the impact of *variability* on the SRAM sense amplifier *offset-voltage*, while considering *both process and run-time variation*. To the best knowledge of the authors, this is the first work to determine the SA offset in the presence of all kind of variability. The used method is accurate in the sense that it uses the *Atomic Model* for aging (which is a calibrated model [9,10]) and considers the *workload dependency* (as the aging variations are strongly workload dependent [11,12]). Guaranteeing a resilient SA requires not only a

978-1-4673-9039-2/16 $31.00 © 2016 IEEE
DOI 10.1109/ISVLSI.2016.30

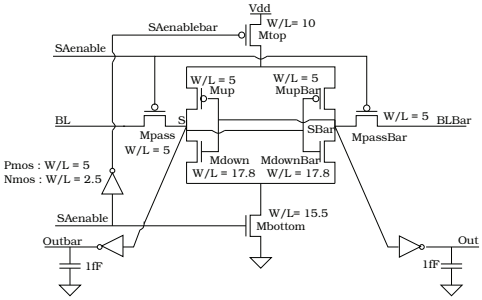


Fig. 1. Standard latch-type Sense Amplifier

correct sensing delay, but also an appropriate offset-voltage during the memory operational lifetime. The results show incorporating the run-time variability due to the aging in the estimation worsens the impact on the offset-voltage with a factor 2 at least.

The rest of the paper is organized as follows. Section II provides a background w.r.t. the targeted standard latch-type sense amplifier and variability sources. Section III provides the proposed methodology for offset voltage quantification. Section IV analyzes the results. Finally, Section V concludes the paper.

II. BACKGROUND

First the standard latch-type sense amplifier is presented, and thereafter, the sources of variability considered in this paper are discussed.

A. Sense Amplifier

Figure 1 depicts the structure of the Standard latch-type Sense Amplifier; it is responsible for the amplification of a small voltage difference between BL and BLBar during read operations. The operation of the sense amplifier consists of two phases. In the first phase, when SAenable is low, the access transistors Mpass and MpassBar connect to the BL (BLBar) with the internal nodes S (Sbar). In the second phase, when SAenable is high, the pass transistors disconnect the BL (BLBar) input from the internal nodes. The cross coupled inverters get their current from Mtop and Mbottom and subsequently amplify the difference between S and Sbar and produce digital outputs on Out and Outbar. The positive feedback loop ensures low amplification time and produces the read value at its output.

B. Variation sources

The four sources of variability investigated in this paper are briefly described next.

Process variations: These affect the circuit at time $t = 0$ and consist of variations in several parameters including effective channel length (L), oxide thickness (t_{ox}), dopant concentration (N_a), and transistor width (W). There are two sources of variation, i.e., systematic and random variation. We focus only on random process variation. It can be described by a probability distribution and can be modeled by V_{th} variation. The standard deviation of the V_{th} shift is given by:

$$\sigma_{V_{TH0}} = \frac{A_{\Delta V_{TH}}}{\sqrt{2WL}} \quad (1)$$

where $A_{\Delta V_{TH}}$ is the Pelgrom's constant [14], W and L the transistor width and length, respectively.

Temperature variation: They impact the operating condition of MOS transistors. The dependence of threshold voltage and temperature is given by [24].

$$V_{th} = C_{vt} - \frac{Q_{ss}}{C_o} + \Phi_{ms} \quad (2)$$

where C_{vt} is a constant that represents the fermi potential, surface charge Q_{ss} at the Si-SiO_2 interface, gate oxide capacitance C_o and work function difference Φ_{ms} is a function of the temperature [24].

$$\Phi_{ms} = -0.61 - \Phi_F(T) \quad (3)$$

Here Φ_F(T) is Fermi potential. Expression (3) shows that work function difference reduces with respect to increase in temperature and therefore leads to a threshold voltage decrement.

Supply voltage variation: Supply voltage fluctuations affect the operating speed of MOS transistors. The variation in switching activity across the die/circuitry leads to an uneven power/current demand and may lead to logic failures [13]. Furthermore, transistor subthreshold leakage variations impact the uneven distribution of supply voltage across the circuitry as well [13]. Hence, reducing the supply voltage degrades the performance of the circuit/transistors and raising supply voltage compensates/enhances the performance and significantly reduces circuit failure rates as a result of variability [13].

Aging Variations: There are different aging mechanisms such as Bias Temperature Instability (BTI) [4], Hot Carrier Injection [15], and Time Dependent Dielectric Breakdown [16]; BTI is considered to be the most important of them; therefore it is the focus of this paper. BTI has two main components i.e., Negative (BTI) and Positive (BTI). Atomistic model is proposed to accurately model BTI [9]; it induces threshold voltage variation for time $t > 0$ and is based on the capture and emission of single traps during stress and relaxation phases of NBTI/PBTI, respectively. The threshold voltage shift of the device ΔV_{th} is the accumulated result of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by [18] as:

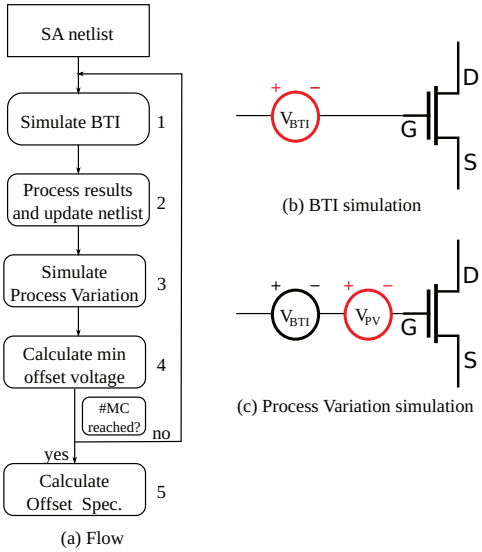


Fig. 2. Offset voltage specification flow.

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c+\tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{STRESS}\right]\right\} \quad (4)$$

$$P_E(t_{RELAX}) = \frac{\tau_e}{\tau_c+\tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{RELAX}\right]\right\} \quad (5)$$

where τ_c and τ_e are the mean capture and emission time constants, and t_{STRESS} and t_{RELAX} are the stress and relaxation periods, respectively. Furthermore, BTI induced V_{th} is an integral function of capture emission time map, workloads, duty factor and transistor dimensions, which gives the mean number of available traps in each device [11]; the model also incorporates the temperature impact [9,10].

III. Proposed Methodology

Figure 2(a) depicts a generic flow to determine the offset voltage specification. It consists of five steps. The first four steps are based on 400 Monte Carlo simulations and are repeated for each experiment. Step 5, the offset specification is based on the entire population. Each of the steps is described next in more details.

1. Simulate BTI: We use the approach in [19] to perform the BTI simulations; they are based on the model described in Section II. The simulations are controlled by a Perl script that generates an initial instance of the BTI augmented SRAM sense amplifier design, based on transistor dimensions, stress time, duty factor and frequency. Every generated instance has a distinct number of traps (with their unique timing constants) in each transistor, and are incorporated in a Verilog-A module of the SA netlist. The module responds to the every individual trap, and alters the transistors concerned parameters. All these parameters can be effectively modeled by a voltage source (the so called BTI-induced threshold voltage) as shown in Figure 2(b). The severity of the BTI impact depends on the workload, temperature, etc. The workload sequence is assumed to be replicated once completed until the age time is reached. Based on the workload, we extract individual duty factors for each transistor based on the waveforms and workload sequence. This enhances the accuracy of our simulation results.

2. Process results and update netlist: In this step, the BTI induced V_{th} shifts of each individual transistor are extracted from the previous step and injected as a voltage source to the netlist. Note that due to the stochastic nature of BTI, each instantiation will have different threshold voltage shifts.

3. Simulate process variations: The next step simulates the process variations as shown in Figure 2(c). Here we follow the same approach as in [6] where the time zero V_{th} variations are modeled by voltage sources. We use the build-in Monte Carlo simulations in Spectre [20] to create a normal distribution for each transistor with a zero mean and a sigma expressed by Equation 1.

4. Calculate minimum offset voltage: The SA offset voltage is crucial for the correct operation of any SA design. The minimum offset voltage of a specific SA instance is the voltage difference between SA inputs (Bit lines) where the cross-coupled inverters of the SA remain in their metastable point. This minimum offset voltage is determined by applying a binary search on the input voltage and is affected by process, temperature, voltage and aging variations.

5. Calculate offset specification: The offset specification of the SA is calculated based on 400 Monte Carlo samples and depends on the desired failure rate or yield. In a good SA design, the offset voltage has a nearly normal distribution and therefore the relation between this distribution and failure rate can be summarized as:

$$\int_{V_{in}=-V_{Offset}}^{V_{Offset}} \mathcal{N}(\mu_{MC},\sigma_{MC}) = 1 - f_r \quad (6)$$

In this equation, V_{in} presents the input voltage of the SA, V_{offset} the offset voltage specification, $\mathcal{N}$ a normal distribution of the offset voltages obtained from step 4, μ_{MC} and σ_{MC} their mean and standard deviation, and f_r the failure rate. The equation states that all SA instantiations that require an offset outside the range [-V_{offset} , +V_{offset}] result in failures. The objective is to find the SA offset specification V_{Offset}. At time t=0, this equation can be solved as follows [21]:

$$V_{Offset} = |normvinv(\tfrac{f_r}{2}, \mu_{MC}=0, \sigma_{MC})| = f(f_r)\cdot\sigma_{MC} \quad (7)$$

In this equation, $norminv$ presents the normal inverse cumulative distribution function which provides the offset voltage for a given f_r, μ_{MC}=0 and σ_{MC}. The equation can be simplified as shown at right-hand side; here f is a function that presents a constant depending on f_r. In this work we assume a constant failure f_r=10^{-9} leading to $f(f_r)$=6.1; therefore, for time t=0 we obtain $V_{Offset} = 6.1 \cdot \sigma_{MC}$. The right-hand side of Equation 7 is only valid when $\mu_{MC} = 0$. However, depending on the workload for time t>0 aging can shift this distribution (i.e., have a non-zero mean); this invalidates Equation 7. As a consequence, we determine V_{Offset} directly from Equation 6 by solving this equation numerically.

IV. SIMULATION RESULTS

In this section, we present the performed experiment, and the obtained results.

A. Experiments Performed

In order to investigate the impact of both zero-time variability and run-time variability, we performed two sets of experiments.

- **Process variation (PV):** In this experiment the impact of process variation is analyzed, while considering first the voltage variations (i.e., -10% V_{dd}, nom. V_{dd}=1.0V, and +10% V_{dd}) and thereafter the temperature variations (i.e., 298K, 348K, and 398K).
- **Process and aging variation:** In this experiment, we analyzed the combined effect of both time-zero and aging variability; also here the experiments are performed for different voltages and temperatures. Moreover, all these experiment are done while considering three different workloads in order to show the dependency on the run-time applications. We assume that 80% of the executed instructions (e.g., by a CPU) are read instructions (which activate the SA). In addition, we define three sequences to present the 80% of the reads: {r0} (all the reads are 0), {r1} (all the reads are 1) and {r0r1} (50% {r0} and 50% {r1}). Note that the aging variability is workload dependent, while PV is not.

B. Simulation Results

Process variability: The results of the process variation for different voltages and temperatures are explained next.

- **Voltage Dependency:** Figure 3 shows the offset voltage distribution for the three supply voltages at nominal temperature at time-zero. The figure shows that the offset distributions are similar and that the impact of voltage variations is *marginal* (less than 5%). For example, w.r.t. nominal V_{dd} a reduction of 4.2% is observed for +10% V_{dd}, while an increase of 2.5% for -10% V_{dd}.

- **Temperature Dependency:** Figure 4 shows the offset voltage distribution for the three temperatures at nominal voltage at time-zero. The figure shows that the offset voltage distributions are nearly the same; hence the impact of temperature variations on the voltage offset is *marginal*. For instance, the offset specification is 90.2mV at 298K while 93.6mV at 398K (an increase of only 3.8%). Note that the mean in Figs. 3 and 4 should ideally be zero; a small error (worst case of 0.19mV) occurs due to the Monte Carlo simulations.

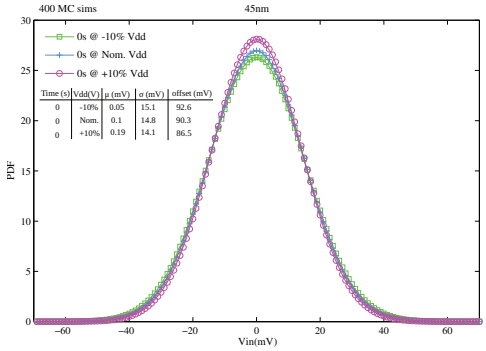


Fig. 3. Supply voltage impact at time-zero.

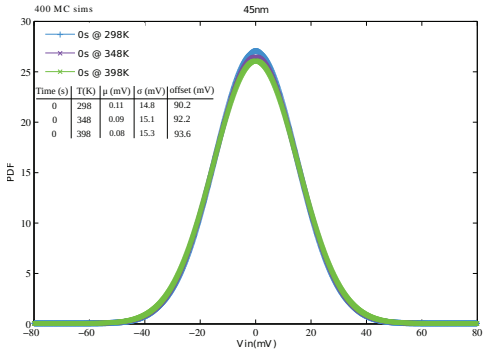


Fig. 4. Temperature impact at time-zero.

Process and aging variability: The results of the voltage and temperature experiments done for both process and aging variations are provided next.

- **Voltage Dependency:** Fig. 5 shows the voltage dependency of the offset voltage for different workloads at 298K. In addition to the baseline plot (time-zero variability), the figure depicts 6 plots: 2V_{dd}'s (-10% and +10%) × 3 workloads; the plots for nominal V_{dd} are not included for clarity. However they are added to Table I, which presents the information of Fig. 5 in another format; the

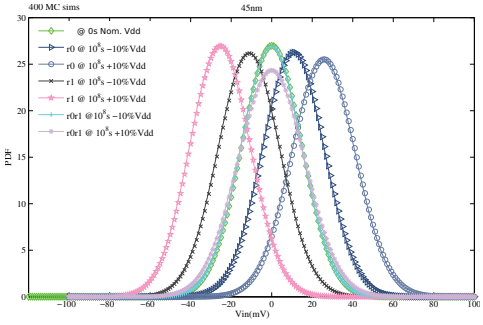


Fig. 5. Voltage impact at run-time.

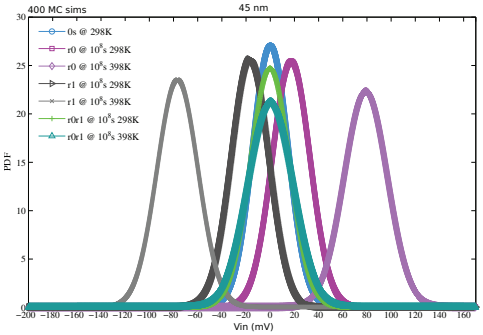


Fig. 6. Temperature variation at run-time.

TABLE I
VOLTAGE IMPACT AT RUN-TIME.

Aging(s)	Workload	Vdd. (V)	μ (mV)	σ (mV)	offset (mV)
0	–	Nom.	0.1	14.8	90.3
10^8	{r0}	-10%	11	15.2	102.0
10^8	{r0}	Nom.	17.3	15.7	111.6
10^8	{r0}	$+10\%$	25.8	15.6	119.5
10^8	{r1}	-10%	-11	15.2	102.4
10^8	{r1}	Nom.	-17.2	15.6	110.6
10^8	{r1}	$+10\%$	-25.3	14.8	114.0
10^8	{r0r1}	-10%	-0.09	14.8	90.6
10^8	{r0r1}	Nom.	-0.2	16.2	98.8
10^8	{r0r1}	$+10\%$	-0.001	16.4	100.0

TABLE II
TEMPERATURE VARIATION AT RUN-TIME.

Aging(s)	Workload	Temp. (K)	μ (mV)	σ (mV)	offset (mV)
0	–	298	0.1	14.8	90.3
10^8	{r0}	298	17.3	15.7	111.5
10^8	{r0}	348	45.0	16.8	145.6
10^8	{r0}	398	79.1	17.9	186.5
10^8	{r1}	298	-17.2	15.6	110.6
10^8	{r1}	348	-44.2	16.3	142.0
10^8	{r1}	398	-76.8	17.0	178.6
10^8	{r0r1}	298	-0.2	16.2	98.8
10^8	{r0r1}	348	-0.02	17.5	107.1
10^8	{r0r1}	398	0.2	18.8	114.8

μ, σ and the offset voltage are also included. Note that the offset voltage distribution for {r0r1} at -10%V_{dd} is almost the same as the baseline, hence they are coincident. The figure shows that depending on the workload, the offset distribution may shift to the left or right; workload {r1} causes a shift to the left, {r0} to the right, while {r0r1} has a nearly zero mean (slight error due to Monte Carlo simulations). This can be explained by the way the workload stresses the SA devices (see Fig. 1); e.g., {r0} stress the transistors Mdown and MupBar of the cross-coupled inverters all the time, {r1} stresses the other devices of the inverters, while {r0r1} balances the stress on all devices.

Moreover, the figure shows that a voltage increase severes the shift as it accelerates the BTI mechanism. Note that although at time 0 the voltage impact is marginal (see Fig. 3), the accelerated aging due to voltage is more significant. For example, for {r0} at nominal V_{dd}, a 17.3mV shift is observed, while this is 25.8mV for +10% V_{dd} (see Table I). A second observation from the figure and table reveals that the σ *increases* with aging, even if balanced workload {r0r1} is applied. Both the average shift and increase in σ lead to *much higher offset requirement*. For example, at nominal V_{dd}, the required offset voltage is only 90.3mV when zero-time variability is considered; however, this is 111.6mV (which is an increase of 24%) when {r0} is applied and run-time variability is considered. It is worth noting that applying a balanced workload {r0r1} results in an offset specification of 98.8mV (which is only an increase of 9.4%); this indicates the importance of balanced workload for optimal designs.

- **Temperature Dependency:** Fig. 6 shows the temperature dependency of the offset voltage at nominal V_{dd}. In addition to the baseline plot (time-zero variability), Fig. 6 depicts 6 plots: 2 temperatures (298K and 398K) $\times$ 3 workloads; the plots for temperature 348K are not included for clarity. However, they are included in Table II. The figure shows similar trends as in Fig. 5, but the impact is more severe.
  The offset specification is strongly dependent on both the workload and temperature. The shift for *unbalanced workload* is *significant*, while this is very marginal for balanced workload. The higher the temperature, the larger the shift. E.g., at T = 398K, the {r0} causes an offset voltage shift of 79mV! This is 76% more than the shift at T = 348K. Another important observation is that irrespective of the workload, the required offset voltage is higher when considering run-time variation. Obviously,

the offset voltage increase is much higher for unbalanced workloads and higher temperatures.
Moreover, the figure and table show that σ increases with temperature for all workload. Both the mean shift and standard deviation lead to *much higher offset voltage specification* irrespective of the workload; the higher the temperature, the higher the required offset voltage. E.g., the offset specification at T = 298K is just 90.3mV when only zero-time variability is considered, while this is 186.5mV (i.e., an increment of 106.5%!) at 398K for workload {r0}. Note that applying a balanced workload minimizes the impact; e.g., an offset specification of 114.8mV (this is up to 27.1%) is obtained at 398K when {r0r1} is applied.

C. Discussion

The proposed methodology for offset voltage quantification is unique not only in the sense that it uses an accurate BTI model and involves the workload dependency, but also because it involves both the zero-time variability as well as run-time variability. The obtained results clearly show that using only environmental run-time variability while considering zero-time variability analysis is not accurate enough. The offset voltage difference between time-zero and run-time variability can be as big as a factor of two.

Moreover, the dependency of offset voltage on workload (application) has been shown to be significant. Applying balanced workload results in reduced impact. Hence, thinking about incorporating some features in the circuits to internally create a balance workload during the lifetime of the application is important for optimal and reliable designs. Schemes such as bit-flipping [23] can be useful.

Finally, it is worth noting that the presented methodology of Figure 2 can be extended to any digital circuit, as long as there is a clear metric (such as offset specification) to be evaluated. For example, the critical paths in a pipeline stage can be evaluated based on the path delay metric.

V. CONCLUSION

This paper used an accurate offset voltage quantification method for the sense amplifier, considering both time-zero and run-time variability; the method takes into consideration process, voltage, and temperature variations, as well as degradation due to aging. The method incorporated also the workload dependency. The results showed a marginal impact of temperature and voltage on offset specification when considering process variability only, while this becomes more significant (up to 2×) when adding the run-time variability due to aging. Hence, ignoring the run-time variability is not suitable for a reliable and robust design. The proposed method gives designers a better way of determining the offset voltage specification which reduces the probability of having over- or under-design margins.

REFERENCES

[1] ITRS, “International Technology Roadmap for Semiconductor 2004” ”www.itrs.net/common/2004 update/2004update.htm.”.

[2] S. Borkar, et al “Micro architecture and Design Challenges for Giga scale Integration”, *Pro. of 37th IEEE International Symposium on Microarchitecture*, 2004.

[3] S. Hamdioui et al., “Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes”, *Design, Automation and Test in Europe*, 2013.

[4] B. Kaczer, et al., ”Atomistic Approach to Variability of Bias Temperature Instability in Circuit Simulation”, Proc. of International Reliability Physics Symposium., April, 2011.

[5] K. Jeong, et al., ”Impact of guardband reduction on design outcomes: A quantitative approach”, IEEE Trans. Semicond. Manuf., vol. 22, no. 4, pp.552-565, Nov. 2009.

[6] S. Cosemans, et al., ”A 3.6pJ/access 480MHz, 128Kbit on-Chip SRAM with 850MHz boost mode in 90nm CMOS with tunable sense amplifiers to cope with variability”, *ESSCIRC*, pp.278-281, Sep. 2008.

[7] M. H. Abu-Rahma, et al., ”Characterization of SRAM sense amplifier input offset for yield prediction in 28nm CMOS”, *IEEE CICC*, pp.1-4, Sep. 2011.

[8] J. Vollrath., ”Signal margin analysis for DRAM sense amplifiers”, *1st IEEE workshop on Electronic Design, Test and Applications*, pp.123-127, 2002.

[9] B. Kaczer, T. Grasser, P. J. Roussel, et al., “Origin of NBTI variability in deeply scaled pFETs”, *IEEE International Reliability Physics Symposium*, 2010.

[10] T. Grasser, et al., ”Analytic modeling of the bias temperature instability using capture/emission time maps”, *IEEE International Electron Devices Meeting*, pp. 1-4, 2011.

[11] P. Weckx, et al., “Defect-based Methodology for Workload-dependent Circuit Lifetime Projections-Application to SRAM”, *IEEE International Reliability Physics Symposium*, April 2013.

[12] D. Rodopoulos, et al., “Time and Workload Dependent Device Variability in Circuit Simulations“ *Proc. Intl. Conf on IC Design and Technology*, pp: 1-4, 2011.

[13] S. Sapatnekar, et al., ”Overcoming Variatins in Nano-scale Technologies”, IEEE Transaction on Emerging and Selected Topics in Circuits and Systems, pp: 5-18, 2011.

[14] M. J. M. Pelgrom, A. C. J. Duinmaijer, et al, “Matching properties of MOS transistors”, *IEEE J. Solid-State Circuits* , Vol. 24, no. 5, pp. 1433-1439, Oct. 1989.

[15] M. Kamal, et al., ”An efficient reliability simulation flow for evaluating the hot carrier injection effect in CMOS VLSI circuits”, IEEE ICCD, pp.352-357, 2012.

[16] M. Choudhury, et al., ”Analytical model for TDDB-based performance degradation in combinational logic”, Design, Automation and Test in Europe, pp.423-428, Mar. 2010.

[17] H. Kukner et al, “Comparison of Reaction-Diffusion and Atomistic Trap-based BTI Models for Logic Gates”, *IEEE trans. on device and materials reliability*, 2014.

[18] M. Toledano-Luque, et al., “Response of a single trap to AC Negative Bias Temperature Stress”, *IEEE International Reliability Physics Symposium*, 2011.

[19] H. Kukner, P. Weckx, J. Franco, M. Toledano-Luque, Moonju Cho, et al., ”Scaling of BTI reliability in presence of Time-zero Variability ”, *IEEE International Reliability Physics Symposium*, 2014.

[20] Cadence., ”Spectre Circuit Simulator datasheet”, from http://cadence.com.

[21] S. Cosemans, “Variability-aware design of low power SRAM memories”, *Ph.D Thesis Katholieke Universiteit Leuven*, 2009.

[22] J. Wang, et al., ”Statistical modeling for the minimum standby supply voltage of a full SRAM array”, *European Solid State Circuit Conference*, pp.400-403, Sep. 2007.

[23] A. Gebregiorgis, et al., ”Aging Mitigation in Memory Arrays Using Self-controlled Bit-flipping Technique”, *Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 231-236, Jan. 2015.

[24] R. Wang, J. Dunkley, T. A. DeMassa, et al., ”Threshold Voltage Variation with Temperature in MOS Transistors”, *IEEE Transaction on Electron Devices*, pp: 386- 388, 1971.

2

Sense Amplifier Offset Voltage Analysis for both Time-zero and Time-dependent Variability

INNOCENT AGBO, Delft University of Technology, the Netherlands
MOTTAQIALLAH TAOUIL, Delft University of Technology, the Netherlands
DANIEL KRAAK, Delft University of Technology, the Netherlands
SAID HAMDIOUI, Delft University of Technology, the Netherlands
PIETER WECKX, imec vzw, Belgium
STEFAN COSEMANS, imec vzw, Belgium
PRAVEEN RAGHAVAN, imec vzw, Belgium
FRANCKY CATTHOOR, imec vzw and KU Leuven, Belgium
WIM DEHAENE, Katholieke Universiteit, Belgium

This paper presents an accurate technique to extensively analyze the impact of time-zero (i.e., global and local variation) and time-dependent (i.e., voltage, temperature, workload, and aging) variation on the offset voltage specification of a memory sense amplifier design using 45 nm predictive technology model (PTM) high performance library. The results show that increasing the supply voltage both for time-zero and time-dependent reduces the offset voltage specification marginally, irrespective of the process corners. In contrast, the offset voltage specification is very sensitive to the temperature and the workload, i.e., the applied voltage patterns. The results also show that a balanced workload results in a significantly lower offset voltage specification. The above results can be used to estimate the required offset voltage accurately for a given lifetime, and operational conditions such as workload, temperature, and voltage; hence, enable the designer to take appropriate measures for a high quality, robust, optimal and reliable design.

Additional Key Words and Phrases: Offset voltage, zero-time variability, time-dependent variability, SRAM sense amplifier (SA), Reliability

ACM Reference Format:
Innocent Agbo, Mottaqiallah Taouil, Daniel Kraak, Said Hamdioui, Pieter Weckx, Stefan Cosemans, Praveen Raghavan, Francky Catthoor, and Wim Dehaene. YYYY. Sense Amplifier Offset Voltage Analysis for both Time-zero and Time-dependent Variability. *ACM Trans. Des. Autom. Electron. Syst.* 9, 4, Article 39 (March YYYY), 20 pages. https://doi.org/0000001.0000001

Authors' addresses: Innocent Agbo, Delft University of Technology, Quantum & Computer Engineering, Delft, South Holland, 2628 CD, the Netherlands, I.O.Agbo@tudelft.nl; Mottaqiallah Taouil, Delft University of Technology, Quantum & Computer Engineering, Delft, South Holland, 2628 CD, the Netherlands, M.Taouil@tudelft.nl; Daniel Kraak, Delft University of Technology, Quantum & Computer Engineering, Delft, South Holland, 2628 CD, the Netherlands, D.H.P.Kraak@tudelft.nl; Said Hamdioui, Delft University of Technology, Quantum & Computer Engineering, Delft, South Holland, 2628 CD, the Netherlands, S.Hamdioui@tudelft.nl; Pieter Weckx, imec vzw, ESAT, Kapeldreef 75, Leuven, B-300001, Belgium, Pieter.Weckx@imec.be; Stefan Cosemans, imec vzw, ESAT, Kapeldreef 75, Leuven, B-300001, Belgium, StefanCosemans.ext@imec.be; Praveen Raghavan, imec vzw, ESAT, Kapeldreef 75, Leuven, B-300001, Belgium, Ragha@imec.be; Francky Catthoor, imec vzw and KU Leuven, ESAT, Kapeldreef 75, Leuven, B-300001, Belgium, Francky.Catthoor@imec.be; Wim Dehaene, Katholieke Universiteit, ESAT, Kapeldreef 75, Leuven, B-300001, Belgium, wim.dehaene@esat.kuleuven.be.

39

1084-4309/YYYY/3-ART39 $15.00
https://doi.org/0000001.0000001

1 INTRODUCTION

In the past decades, CMOS technology has undergone assertive down-scaling that causes significant bottleneck on the reliability of devices [ITRS 2005]. Nowadays, the unreliability sources of nano-scaled technologies are mostly induced by manufacturing or time-dependent variability [ITRS 2005]. During the manufacturing process, devices are affected by process variation, which changes the properties of the manufactured devices from the targeted ones. Consequently, the *process* or *time-zero variation* [Borkar 2004; Wang et al. 2007] is referred to as similar manufactured devices end up having various characteristics.

They can be further classified into global (i.e., different process corners) and local (i.e., mismatch) variations.

In addition, *time-dependent variations* result in the device properties to vary and/or degrade during their operational lifetime. Such variations are mostly as a result of *environmental variations* for instance, supply voltage variations and temperature fluctuations [Borkar 1999; Hamdioui et al. 2013; Marinissen et al. 2016; Ray et al. 2016], and *temporal* or *aging variations* for example, *Bias Temperature Instability (BTI)*; both depict an extremely growing effect with CMOS scaling [Alam and Mahapatra 2005; Kaczer et al. 2011; Zafar et al. 2006]. All these variations result in the devices to behave in a different manner than proposed, which may result in the devices (or circuits) to fail if adequate measures are not taken. Designers normally utilize a conservative guard-band and use extra design margins [Jeong et al. 2009] to assure the proper operation for the worst-case variations until the aimed circuit lifetime. Nevertheless, a pessimistic guard-band guides to either yield or performance loss, whilst an optimistic guard-band enhances the test escapes and in-field failures. Therefore, it is crucial to consider *all kinds of* variations for both manufacturing and time-dependent to estimate their impact accurately to prevent an overly pessimistic design. Guaranteeing a resilient Sense Amplifier (SA) requires not only a correct sensing delay, but also an appropriate offset-voltage during the memory operational lifetime.

Much work has been done on the impact of reliability on *Static Random Access Memory (SRAM)* cell array but not much work has been investigated for the offset voltage characterization in SAs. Kumar *et al.* [Kumar et al. 2006] and Andrew [Carlson 2007] investigated the impact of Negative Bias Temperate Stability (NBTI) on the read stability and the Static Noise Margin (SNM) of SRAM cells. Khan *et al.* [Khan et al. 2014] analysed the impact of BTI on FinFET based memory cells for different SRAM designs using SNM, Read Noise Margin (RNM) and Write Triple Point (WTP) as metrics. In [Cosemans et al. 2008], the authors investigated a tunable SA to deal with with-in die variations; the authors predicted the offset voltage at design time built on process variations. In [Abu-Rahma et al. 2011], the authors distinguished the SA Input Offset by a physical monitoring circuit (implemented in silicon) in order to estimate the yield. In [Vollrath 2002], the authors investigated a technique to ascertain the signal margins for *Dynamic Random Access Memory (DRAM)* SAs built on offset distribution quantification. The above prior work only focused on time-zero variation. It is crucial to use the right techniques to quantify the offset voltage. Hence, in our previous work [Agbo et al. 2016], we presented an accurate technique to estimate the impact of both time-zero and time-dependent variability on the offset voltage of sense amplifiers, while considering various workloads and process voltage temperature (PVT) variations. However, the impact of global process variation has not been considered. In addition, both SA's sensitivity and reliability failure analysis have not been explored yet.

In this respect, the main contributions of the paper are:

- Sensitivity analysis of the SA.

- Investigation of the SA offset voltage specification while considering five global process corners at time-zero.
- Thorough voltage, temperature and workload analysis of process and time-dependent variation on the offset voltage specification while considering different process corners.
- Failure rate analysis for the offset voltage specification at *Nominal* process corner, while considering different workloads, supply voltages, and temperatures.

The results show that by incorporating time-zero and time-dependent variability, the offset voltage specification worsens for the *Fast-Slow (FS)* corner with a factor up to 3× as compared to the *Slow-Fast (SF)* corner. Hence, it is of great importance to take this into consideration when designing robust and reliable memory systems.

The rest of the paper is organized as follows. Section 2 provides a background with respect to the targeted standard latch-type sense amplifier and variability sources. Section 3 discusses the proposed methodology for offset voltage quantification. Section 4 presents the experiments performed. Section 5 analyzes the sensitivity and time-zero variability results. Section 6 analyzes both the time-zero and time-dependent variability results. Section 7 analyzes the reliability failure rate results. Section 8 discusses the crucial messages of the paper. Finally, Section 9 concludes the paper.

2 BACKGROUND

The standard latch-type sense amplifier which is used in this work is first presented. Thereafter, the variability sources studied in this paper are explained.

2.1 Sense Amplifier

Fig. 1 shows the structure of Standard Latch-Type Sense Amplifier (SLTSA) [Cosemans 2009], it is in-charge for the amplification of a small voltage change between BL and BLBar during read operations. It produces amplified signals on the output (i.e., *Out* and *Outbar*). The circuit consists of pull-up transistors (i.e., Mup and MupBar), pull-down transistors (i.e., Mdown and MdownBar), two access transistors (i.e., Mpass and MpassBar), two current source transistors (i.e., Mtop and Mbottom), and two inverters at the output of the Sense Amplifier each with a load capacitance of 1fF. The internal cross coupled inverters consist of two pull-up transistors Mup and MupBar and two pull-down transistors Mdown and MdownBar. They have BL and BLBar as inputs and produce amplified outputs (i.e., *Out* and *Outbar*).

The operation of the sense amplifier comprises of two stages. In the first stage, while SAenable is low, the access transistor Mpass connects BL with the internal node *S* while the access transistor MpassBar connects BLBar with the internal node *SBar*. In this stage, Mtop and Mbottom transistors are turned off. In the second stage, while SAenable is high, the access transistors disconnect the BL(BLBar) input from the internal nodes. The cross coupled inverters obtain their current from Mtop and Mbottom and afterward amplify the change in voltage between *S* and *Sbar*. *S(SBar)* node is actively pulled down when *SBar(S)* exceeds the threshold voltage of Mdown. The positive feedback loop ensures low amplification time and produces the read value at its output. All current paths are disabled when *S(SBar)* is at 0V and *SBar(S)* is at V_{dd} or vice versa. This process is repeated for each read operation.

2.2 Variation sources

In this paper we analyzed four sources of variability as explained next.

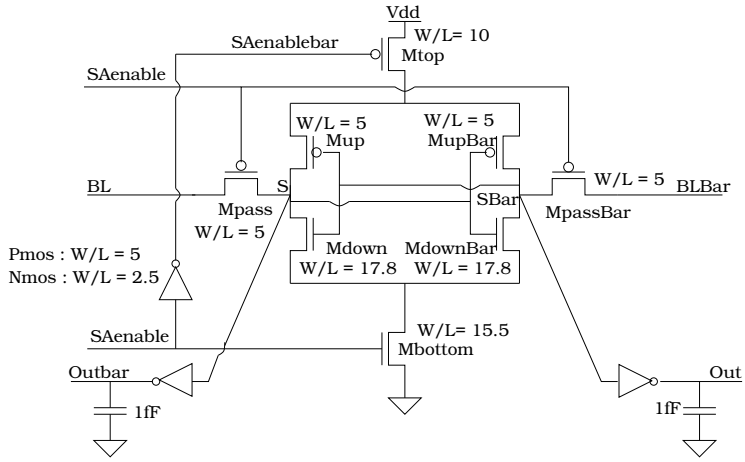


Fig. 1. Standard latch-type Sense Amplifier.

Process variations (PV): These impact the circuit at time $t = 0$ and comprise of variations in various parameters incorporating effective channel length (L), oxide thickness (t_{ox}), dopant concentration (N_a), transistor width (W), and threshold voltage V_t. The variations can be classified into local and global variations. Each is briefly described.

- *Local variation:* Local variation can be modeled by variation in V_{th} with a standard deviation equal to:

$$\sigma_{V_{TH0}} = \frac{A_{\Delta V_{TH}}}{\sqrt{2WL}} \tag{1}$$

  where $A_{\Delta V_{TH}}$ is the Pelgrom's constant [Pelgrom et al. 1989], and W is the transistor width while L is the transistor length.

- *Global variation:* Global process variations are typically a consequence of wafer etching and deviations during the lithography process [Weste and Harris 2010]. These variations are constant over larger areas, and lead to different process corners. In this work we consider the following corners: typical-typical (*TT*), fast-fast (*FF*), slow-slow (*SS*), slow-fast (*SF*) and fast-slow (*FS*). The *TT*, *FF*, and *SS* are uniform corners since they impact both NMOS and PMOS devices uniformly. However, *FS* and *SF* cause an unequal switching in the circuit [Weste and Harris 2010].

Temperature variation: They affect the operating condition of MOS transistors. An increase in temperature causes a threshold voltage reduction (which has a positive effect on delay), and a carrier mobility decrement (which has a negative impact on delay and consequently leakage current increment [Wang et al. 1971]).
The reliance of the threshold voltage on the temperature is stated by [Wang et al. 1971]:

$$V_{th} = C_{vt} - \frac{Q_{ss}}{C_o} + \Phi_{ms} \tag{2}$$

where C_{vt} is a constant that denotes the fermi potential, Q_{ss} denotes the surface charge at the $Si\text{-}SiO_2$ interface, C_o denotes the gate oxide capacitance and Φ_{ms} denotes the change in work function depending on the temperature, T [Wang et al. 1971] as stated next:

$$\Phi_{ms} = -0.61 - \Phi_F(T) \tag{3}$$

Here Φ_F(T) represents the Fermi potential. Expression (3) depicts that the change in work function reduces with respect to a rise in temperature and therefore causes a threshold voltage decrement.

Supply voltage variation: The change in supply voltage impacts the operating speed of MOS transistors. The fluctuation in switching activity across the die/circuitry causes an irregular power/current demand and may cause logic failures [Sapatnekar 2011]. In addition, transistor sub-threshold leakage fluctuations affect the irregular distribution of supply voltage across the circuitry as well [Sapatnekar 2011]. Thus, reducing the supply voltage degrades the performance of the circuit/transistors and raising supply voltage compensates/improves the performance and significantly reduces circuit failure rates due to variability [Sapatnekar 2011].

Aging Variations and modeling: There are various aging techniques such as Bias Temperature Instability (BTI) [Kaczer et al. 2011], Hot Carrier Injection [Kamal et al. 2012], and Time Dependent Dielectric Breakdown [Choudhury et al. 2010]; BTI is assessed to be the most significant of them [Grasser et al. 2011; Kaczer et al. 2010; Rodopoulos et al. 2011; Weckx et al. 2013]; hence it is the focal point of this paper. BTI has two major components i.e., Negative (BTI) and Positive (BTI). Atomistic model is proposed to correctly model BTI [Kaczer et al. 2005, 2010; Kukner et al. 2014a,b]; it impacts threshold voltage fluctuation for time $t > 0$ and is founded on the capture and emission of single traps during stress and relaxation stages of NBTI/PBTI, respectively. The threshold voltage variation of the device ΔV_{th} is the amassed result of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by [Toledano-Luque et al. 2011] as:

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c + \tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e} + \frac{1}{\tau_c})t_{STRESS}\right]\right\} \tag{4}$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c + \tau_e}\left\{1 - exp\left[-(\frac{1}{\tau_e} + \frac{1}{\tau_c})t_{RELAX}\right]\right\} \tag{5}$$

where τ_c is the mean capture time constant and τ_e is the mean emission time constant, while t_{STRESS} is the stress period and t_{RELAX} is relaxation period. Moreover, BTI impacted V_{th} is a combined function of capture emission time map, workloads, duty factor and transistor dimensions, which gives the average number of available traps in each device [Weckx et al. 2013]; the model also consists the temperature impact [Grasser et al. 2011; Kaczer et al. 2010].

3 PROPOSED METHODOLOGY

Fig. 2(a) shows a comprehensive flow to ascertain the offset voltage specification of the SA explained. It comprises of five steps. The first four steps are based on 400 Monte Carlo simulations and are replicated for each experiment. Step 5 ascertains the offset specification based on the whole population. Each of the steps is explained next in more details.

1. Simulate BTI: First the SA netlist is generated using 45-nm PTM high performance library [PTM 2012]. To simulate the BTI, we use the atomistic model described in Section 2. The simulations are regulated by a Perl script that produces different instantiations where the BTI induced V_{th}

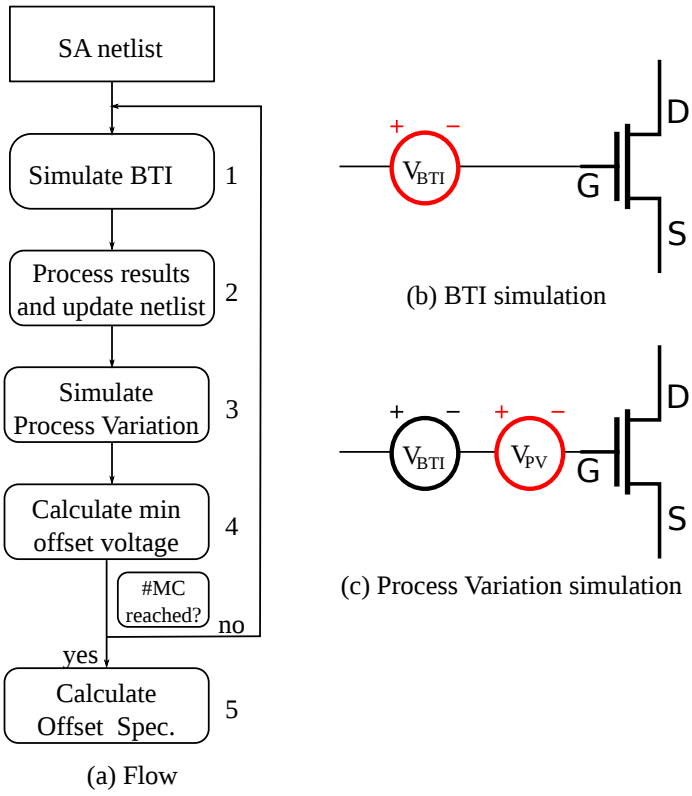


Fig. 2. Offset voltage specification flow.

distribution depends on the transistor dimensions, stress time (3 years), duty factor (which depends on the workloads that will be described in Section 4) and frequency (1.32 GHz) [Borkar 1999]. Every produced instance has a certain number of traps (with their unique timing constants) in each transistor, and are integrated in a Verilog-A module of the SA netlist. The module reacts to every individual trap, and changes the transistors concerned parameters. All these parameters can be effectively modeled by a voltage source (the so called BTI-induced threshold voltage) as depicted in Fig. 2(b). The severity of the BTI effect depends on the workload, temperature, etc. The workload sequence is supposed to be replicated once completed until the age time of three years is attained. Dependent on the workload, we obtain individual duty factors for each transistor depended on the waveforms and workload sequence. This improves the accuracy of our simulation results.

2. Process results and update netlist: In this step, the BTI induced V_{th} shift of each individual transistor is obtained from the earlier step and injected as a voltage source (see Fig. 2(b)) to the netlist. Note that as a result of the stochastic nature of BTI, each instantiation will have varying threshold voltage shift.

3. Simulate process variations: The next step simulates the time zero variations as depicted in Fig. 2(c). Here we follow the same technique as in [Cosemans et al. 2008] where the process V_{th}

variations are modeled by a voltage source, V_{pv}. We utilize the build-in Monte Carlo simulations in Spectre [Cadence 2016] to allocate a V_{pv} based on a normal distribution for each transistor with a zero mean and a sigma related by Eqn. 1.

4. Calculate minimum offset voltage: The SA input voltage is vital for accurate operation of any SA design. We define the minimum *offset voltage* of the SA as the lowest voltage where the SA still successfully completes a read 0 and 1 operations. This voltage can be found for each specific SA instance by looking for the input voltage where the cross-coupled inverters of the SA remain in their metastable point, i.e., where the output starts to change its value. In this work, the minimum offset voltage is determined by applying a binary search on the SA input voltage while considering a sufficient input voltage range (e.g., from -200mV to 200mV). It is worth noting that the offset voltage specification is impacted by process, temperature, voltage, aging variations and workload.

5. Calculate offset specification: The offset specification of the SA is computed based on 400 Monte Carlo samples and depends on the desired failure rate or yield. In a good SA design, the offset voltage has a nearly normal distribution and this means that the relation between this distribution and failure rate can be summarized as [Cosemans 2009]:

$$\int_{V_{in}=-V_{Offset}}^{V_{Offset}} \mathcal{N}(\mu_{MC}, \sigma_{MC}) = 1 - f_r \tag{6}$$

In this equation, V_{in} presents the input voltage of the SA, V_{Offset} means the offset voltage specification, $\mathcal{N}$ is the probability density function (PDF) of the normal distributed offset voltages obtained from step 4; μ_{MC} and σ_{MC} are mean and standard deviation, while f_r the failure rate. The equation states that all SA instantiations that need an offset outside the range [$-V_{Offset}$, $+V_{Offset}$] result in failures. The objective is to find the SA offset specification V_{Offset}. At time t=0, this equation can be solved as follows [Cosemans 2009]:

$$V_{Offset} = |norminv(\frac{f_r}{2}, \mu_{MC} = 0, \sigma_{MC})| = f(f_r) \cdot \sigma_{MC} \tag{7}$$

where *norminv* presents the normal inverse cumulative distribution function providing the offset voltage for a given f_r, μ_{MC}=0 and σ_{MC}. The equation can be simplified as depicted at right-hand side, here f is a function that presents a constant depending on f_r. However, depending on the workload for time t>0 aging can shift this distribution (i.e., have a non-zero mean); this invalidates Equation 7. As a consequence, we determine V_{Offset} directly from Equation 6 by solving this equation numerically. In this work a failure rate f_r=10^{-9} leading to $f(f_r)$=6.1 is assumed. This means that for time t=0 we obtain $V_{Offset} = 6.1 \cdot \sigma_{MC}$. The right-hand side of Equation 7 is only valid when $\mu_{MC} = 0$.

4 EXPERIMENTS PERFORMED

In this work, we performed four sets of experiments for the different process corners while considering the impact of both zero-time variability and time-dependent variability on the offset specification. They are:

- **Sensitivity analysis:** The sensitivity analysis for the SA's offset voltage specification is performed for five process corners (i.e., *Nominal (Nom.), Fast-Fast (FF), Fast-Slow (FS), Slow-Fast (SF), and Slow-Slow (SS)*); this analysis is done without the BTI impact. The target is to

Table 1. Sensitivity analysis at time-zero.

Process corners	σ offset voltage. due to Pass transistors (mV)	due to Pull-down Transistors (mV)	due to Pull-up Transistors (mV)	overall (mV)	offset spec. (mV)
Nom.	0.17	11.16	0.69	15.84	96.06
FF	0.40	11.46	0.80	16.26	96.41
FS	0.28	11.77	0.55	16.69	101.00
SF	0.06	10.68	0.91	15.17	92.30
SS	0.13	10.83	0.60	15.37	94.30

investigate the contribution of each transistor to the offset voltage specification. During this sensitivity analysis, only the results of the pass transistors, pull-down and pull-up transistors are reported. Note that the transistors *Mtop* and *Mbottom* of Fig. 1 do not impact the offset voltage specification, since the transistors *Mtop* and *Mbottom* are common to the cross-coupled inverters of the SA. We have verified this in simulation.

- **Process variation (PV):** The impact of process variation on the offset voltage specification is performed for all process corners, while considering different voltages (i.e., -10% V_{dd}, nominal V_{dd}=1.0V, and +10% V_{dd}) and temperatures (i.e., nominal temperature=298K, 348K, and 398K).
- **Process and aging variation:** We examine the combined effect of both time-zero and time-dependent variability (for 3 years) for the five process corners. The examination is performed while considering varying supply voltage, temperature conditions, and workloads. We suppose that 80% of the executed instructions (e.g., by a CPU) are read instructions (which activate the SA). Furthermore, we define three distinct sequences to introduce the 80% of the reads: {r0} (all the reads are 0), {r1} (all the reads are 1) and {r0r1} (50% {r0} and 50% {r1}). It is worth noting that the aging variability is workload dependent, while PV is not. The explanation for the workload stresses have been given in [Agbo et al. 2016].
- **Reliability failure analysis:** In this experiment, failure rate analysis is performed for nominal the process corner, while considering various supply voltages, temperatures, and workloads.

5 TIME-ZERO VARIABILITY

5.1 A Sensitivity analysis

Table 1 depicts the results of the sense amplifier's sensitivity analysis due to time-zero variability. Note that the sensitivity analysis of the individual transistors (i.e., pass transistors, pull-down and pull-up transistors) is performed using the same approach as explained in Step 5 of Section III, but it considers only process variations.

The table reports the standard deviation of the input voltage of the SA of all the process corners considering PV in individual transistors (column 2 to column 4) and PV in all transistors (column 5). The last column provides the offset voltage specification (derived from Eqn. 6). The table shows that the pull-down transistor is the most dominant contributor to the offset voltage specification, irrespective of the process corner. For example, at nominal process corner, the standard deviation due to the pull-down transistor is 11.16mV, while this does not exceed 0.17mV and 0.69mV for the pass transistor and pull-up transistor, respectively. The pull-down transistors are mostly responsible for discharging the inputs to the SA after being pre-charged; in case these pull-down transistors

Table 2. Voltage and temperature impact at time-zero.

	Values (mV)	V_{dd}			Temp.		
		-10%	Nom.	+10%	298K	348K	398K
Nom. corner	μ	0.09	0.12	0.15	0.12	0.12	0.13
	σ	16.0	15.7	15.4	15.7	16.3	16.9
	Offset	97.51	96.06	94.03	96.06	99.51	103.5
FF corner	μ	0.15	0.19	0.21	0.19	0.18	0.17
	σ	16.2	15.8	15.4	15.8	16.5	17.1
	Offset	99.17	96.41	93.95	96.4	100.6	104.6
FS corner	μ	0.12	0.14	0.16	0.14	0.15	0.18
	σ	16.8	16.5	16.2	16.5	17.4	18.4
	Offset	102.7	101.0	99.0	101.0	106.1	112.7
SF corner	μ	0.06	0.10	0.15	0.10	0.093	0.097
	σ	15.3	15.1	14.7	15.1	15.5	15.9
	Offset	93.6	92.3	90.1	92.3	94.6	96.9
SS corner	μ	0.07	0.08	0.10	0.08	0.10	0.12
	σ	15.5	15.4	15.3	15.4	16.0	16.8
	Offset	94.90	94.29	93.20	94.30	97.70	102.4

are implemented with the fast corner library (i.e., they have a lower V_{th}), the impact of local PV is larger, leading to a larger offset voltage impact. In addition, the table also shows that the shift in offset voltage is the worst (5%) in case of *FS* process corner as compared with nominal case, while the *SF* corner reduces the offset specification voltage with 4.1%.

5.2 Process variability

The results of the impact of PV experiments, varying supply voltages, and temperatures on the voltage offset are discussed next.

Voltage Dependency

Fig. 3 depicts the probability density function (PDF) of the normal distributed offset voltages of three process corners (i.e., *FS*, *Nom.*, and *SF*) for two voltages (i.e., -10%V_{dd} and +10%V_{dd}) at nominal temperature at time-zero. The y-axis represents the PDF, while the x-axis represents the *minimum offset voltage of the SA* denoted as Vin. The plots for the two other process corners (i.e., *FF* and *SS*) are omitted for clearness. Note that the results for *FF* and *SS* are within the boundaries of *FS* (max value) and *SF* (min values). Nevertheless, they are all included in Table 2 which presents all voltage dependency results; the μ, σ and the offset voltage specification are also added. The figure clearly shows that the voltage offset specification dependency on the varying supply voltage at time-zero is marginal, irrespective of the three process corners; the voltage offset increment does not exceed 2.4% in the worst case, which is for *SF* corner. Note that the mean in Table 2 should be zero for all the process corners; however a small error occurs as a result of the Monte Carlo simulations. Note also that relatively the lower the supply voltage, the higher the required voltage offset specification, irrespective of the process corner. For example, at *FF corner* the offset voltage is 99.17mV at -10%V_{dd} while this is 93.95mV at +10%V_{dd}. This can be explained by the impact of PV;

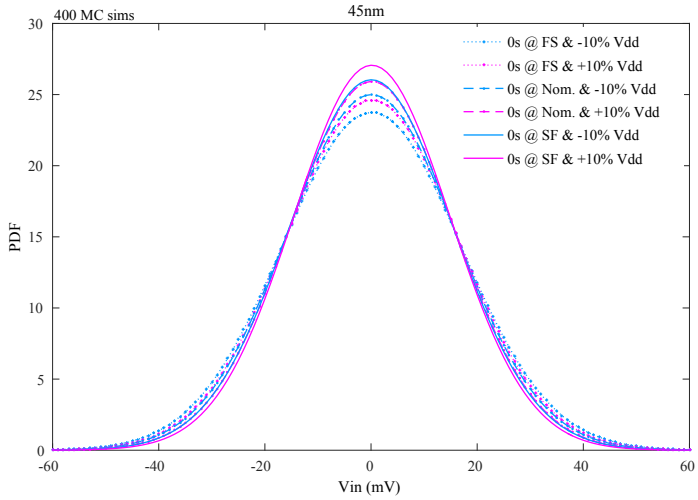


Fig. 3. Supply voltage impact at time-zero for three process corners.

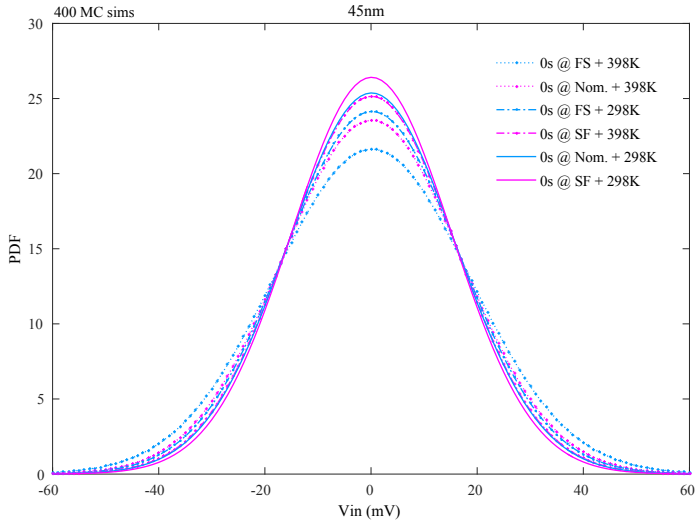


Fig. 4. Temperature impact at time-zero for three process corners.

the lower the V_{dd} the higher the impact of PV.

Temperature Dependency

Fig. 4 depicts the PDF of the offset voltage distribution of three process corners (*FS*, *Nom.*, and *SF*) for two temperatures (298K and 398K) at nominal supply voltage. The plots for the other two process corners (*FF* and *SS*) are omitted for the same reasons as for Fig. 3. Nevertheless, they are all included in Table 2. The figure and table show that the offset voltage specification increases

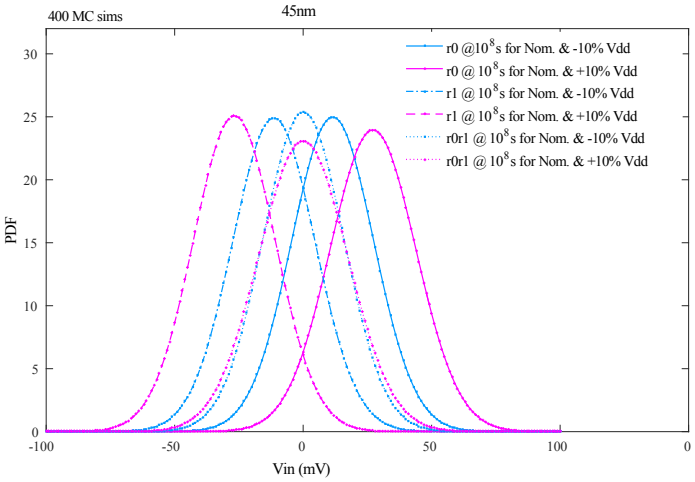


Fig. 5. Voltage impact at time-dependent for *Nominal* corner.

Table 3. Voltage impact at time-dependent.

			Nominal			FF			FS			SF			SS		
Aging (s)	WL	Vdd. (V)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)
0	–	Nom.	0.1	15.7	96.1	0.2	15.8	96.4	0.1	16.5	101.0	0.1	15.1	92.3	0.1	15.4	94.3
10^8	{r0}	−10%	11.5	16.0	107.3	11.9	16.3	109.9	12.0	16.8	112.9	11.2	15.3	103.2	11.1	15.5	104.0
10^8	{r0}	Nom.	18.1	16.6	117.7	18.4	16.8	119.5	28.9	17.8	135.6	17.6	16.0	113.3	17.5	16.2	114.6
10^8	{r0}	+10%	27.2	16.7	127.1	27.4	16.8	127.9	28.3	17.5	133.2	26.5	16.0	122.5	26.6	16.3	124.5
10^8	{r1}	−10%	-11.4	16.0	107.5	-11.7	16.3	109.5	-11.8	16.9	113.0	-11.2	15.4	103.3	-11.0	15.6	104.4
10^8	{r1}	Nom.	-17.9	16.5	117.0	-18.0	16.7	117.9	-28.5	17.0	130.6	-17.5	15.9	112.7	-17.5	16.2	114.5
10^8	{r1}	+10%	-26.8	15.9	122.2	-26.8	16.0	122.5	-27.9	16.7	128.2	-26.1	15.3	117.6	-26.4	15.7	120.6
10^8	{r0r1}	−10%	0.1	15.7	96.0	0.2	15.8	96.6	0.1	16.5	100.9	0.1	15.1	92.3	0.1	15.4	94.0
10^8	{r0r1}	Nom.	-0.2	17.1	104.6	-0.1	17.3	105.5	-0.2	18.0	109.9	-0.2	16.4	100.5	-0.2	16.7	102.2
10^8	{r0r1}	+10%	0.0	17.3	105.6	0.1	17.5	106.7	0.1	18.1	110.9	-0.0	16.6	101.5	-0.0	16.9	103.1

with temperature, reaching 12% for T=398K *FS* (worst case) as compared with T=298K. The figure and table also show that the average, μ=0, irrespective of the process corners. Note that higher temperatures lead to higher σ shifts, irrespective of the process corners; this is caused by the impact of temperature on the V_{th}. V_{th} reduces with temperature; therefore, the impact of PV is bigger at higher temperatures. For example, at *nominal* corner the standard deviation is 15.7mV at 298K and 16.9mV at 398K; an increment of 7.6%. In conclusion, the impact of PV at time zero on the SA voltage offset is marginally dependent on supply voltage (only 2.4% increment), while it is reasonably dependent on the temperature (up to 12% increment).

6 TIME-DEPENDENT VARIABILITY

The same experiments of the previous section will be performed but now considering both time-zero and time-dependent variability. As time-dependent variability is workload dependent, also different workloads will be analyzed.

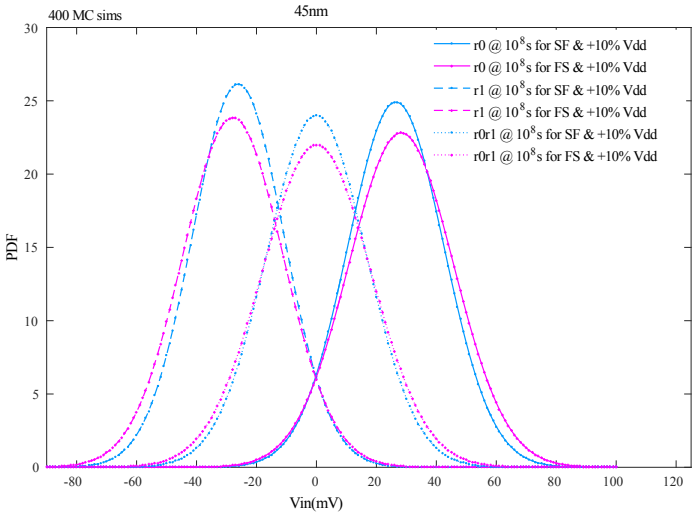


Fig. 6. Voltage impact at time-dependent for both *SF* & *FS* corner at +10% V_{dd}.

6.1 Voltage Dependency

Fig. 5 depicts the PDF of the offset voltage for different supply voltages and workloads using the *Nominal* process corner at 298K. For the sake of clarity, only *Nominal* corner has been provided. The other cases can be found in Table 3. The figure shows 6 plots: $2V_{dd}$'s (-10% and +10%V_{dd}) × 3 workloads (see Section IV). The plots for the nominal V_{dd}'s and for the other process corners are left out for clarity reasons; nevertheless, all the results are listed in Table 3. The figure depicts that unbalanced workloads shift the mean offset voltage distribution to the right (for {r0}) or left (for {r1}), while the mean shift of balanced workloads ({r0r1}) equals zero. Although local PV causes mean shifts, BTI might influence them. Note that in BTI only stressed transistors are impacted and therefore an unbalanced workload will lead to an unbalanced cross coupled inverter pair (see transistors *Mup*, *Mupbar*, *Mdown*, and *Mdownbar* in Fig. 1). This will favor either {r0} or {r1} operations and eventually cause the means to shift.

The figure also shows that this mean offset voltage shift worsens for higher V_{dd}. A V_{dd} increment accelerates the BTI process due to a larger electrical field across the gate. This mean shift eventually impacts the BTI offset voltage specification negatively. For example, in Fig. 5 and Table 3, for {r0} at Nom.V_{dd}, the mean shift is 18.1 mV, while this is 27.2 mV at +10%V_{dd} and 10^8s. The figure and table also show that the standard deviation of the offset voltage shift marginally increases both for unbalanced and balanced workloads. Furthermore, the combined impact of both mean and standard deviation offset voltage shift significantly increases the offset voltage specification for unbalanced workloads (up to 127.1 mV); while this marginally increases for the balanced workloads (up to 105.6 mV).

Table 3 also tabulates the results of the other process corners. The table depicts that the offset voltage specification significantly increases at 10^8s with respect to time-zero at Nom.V_{dd}, irrespective of the process corners. For example, at time-zero and *FF* corner, the offset voltage specification is 96.4mV at time-zero while it increases to 119.5mV at 10^8s, (which is up to 23.6%); similar results

Table 4. Temperature variation at time-dependent.

			Nominal			FF			FS			SF			SS		
Aging (s)	WL	Temp. (K)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)	μ (mV)	σ (mV)	offset (mV)
0	–	298	0.1	15.7	96.1	0.2	15.8	96.4	0.1	16.5	101.0	0.10	15.1	92.3	0.1	15.4	94.3
10^8	{r0}	298	18.1	16.6	117.7	18.4	16.8	119.5	18.8	17.5	123.5	17.6	16.0	113.3	17.5	16.2	114.6
10^8	{r0}	348	47.7	18.1	156.5	49.1	18.6	160.7	50.5	19.5	167.3	46.1	17.2	149.0	46.5	17.7	152.9
10^8	{r0}	398	87.1	20.3	208.9	89.1	20.6	212.6	94.8	22.7	231.1	82.5	18.7	194.5	86.3	20.4	208.6
10^8	{r1}	298	-17.9	16.5	117.0	-18.0	16.7	117.9	-18.6	17.4	122.7	-17.5	15.9	112.7	-17.5	16.2	114.5
10^8	{r1}	348	-46.6	17.5	151.3	-47.6	17.7	153.9	-49.0	18.6	160.4	-45.2	16.6	144.8	-45.4	17.1	147.9
10^8	{r1}	398	-83.0	18.6	194.4	-84.9	18.9	198.2	-89.0	20.1	209.5	-79.4	17.4	183.9	-81.6	18.3	191.5
10^8	{r0r1}	298	-0.2	17.1	104.6	-0.1	17.3	105.5	-0.2	18.0	109.9	-0.2	16.4	100.5	-0.2	16.7	102.2
10^8	{r0r1}	348	0.0	18.8	114.8	0.1	19.2	117.1	0.1	20.1	122.5	-0.0	17.8	109.0	0.0	18.4	112.3
10^8	{r0r1}	398	0.3	20.7	126.6	0.3	21.1	129.1	0.4	22.7	138.5	0.2	19.3	118.1	0.3	20.5	125.5

but with different degradations are obtained for the other corners. For *FS*, this increment is 34.3%, while for *SF* only 11.5%.

The table further reveals that the impact of V_{dd} is marginal on the voltage specification. For example, for the *Nominal* corner at +10%V_{dd} the *mean and sigma* are -26.8mV and 15.9mV for workload {r1} and they are equal to -17.9mV and 16.5mV at Nom.V_{dd}.

Fig. 6 depicts the PDF of the offset voltage for the extreme voltage condition i.e., (+10% V_{dd}) for both *FS* (worst case) and *SF* (best case) process corners, irrespective of the applied workloads. The figure shows that at 10^8s and +10% V_{dd}, the offset voltage distribution is wider for *FS* process corner than for *SF* process corner; overall *FS* has a high offset voltage specification. This is due to an increase in both μ and σ offset shift (see Table 3). For example, the offset voltage shift at {r0} is 122.5mV for *SF* corner while this is only 133.2mV for *FS* corner (which is up to 8.7% higher).

6.2 Temperature Dependency

Fig. 7 depicts the PDF of the offset voltage for various temperatures, and workloads for the *Nominal* process corner at nominal voltage. In addition to the time-zero plot; the figure also depicts six plots with respect to time-dependent variability, two temperatures (298K and 398K) × 3 workloads. The plots for 348K and the other four process corners (i.e., *FF*, *FS*, *SF*, and *SS*) are omitted due to limited space. However, their results are included in Table 4.

Fig. 7 shows that the average offset voltage shift increases significantly at a high temperature (398K) for the unbalanced workloads, while the mean offset shift is zero for the balanced workload. For example, at 10^8s and 298K, the mean offset voltage shift is 18.1mV while this increases to 87.1mV at 398K (this is up to 381.2%!) for the unbalanced workload (results of {r0} are taken). The figure and Table 4 also show that a temperature rise increases the standard deviation, irrespective of the workloads. For example, at 10^8s and for 298K the standard deviation equals 16.6mV while this is 20.3mV at 398K (this increases up to 22.3%) for {r0} workload.

The figure and in particular Table 4 show that the offset voltage specification increases significantly, especially for the unbalanced workload. This can be explained by the increment in σ and μ (for unbalanced workloads) when the temperature increases.

The higher the temperature, the larger the offset voltage specification; e.g., at 398K the {r0} causes an offset voltage specification of 208.9mV for nominal corner, while this is 117.7mV at 298K

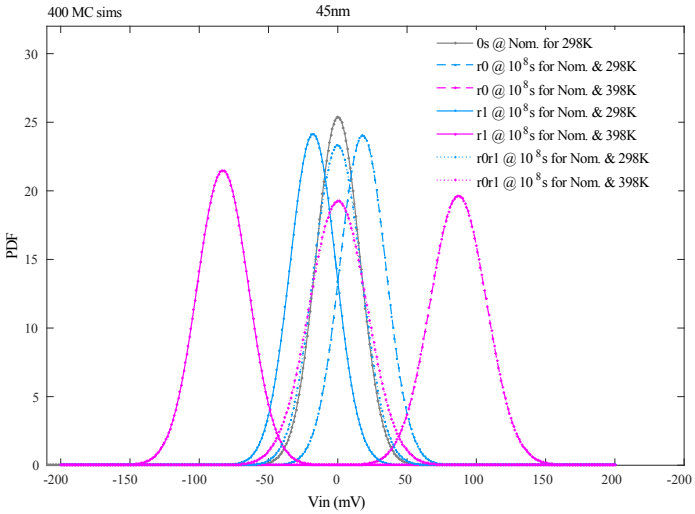


Fig. 7. Temperature impact at time-dependent for *Nominal* corner.

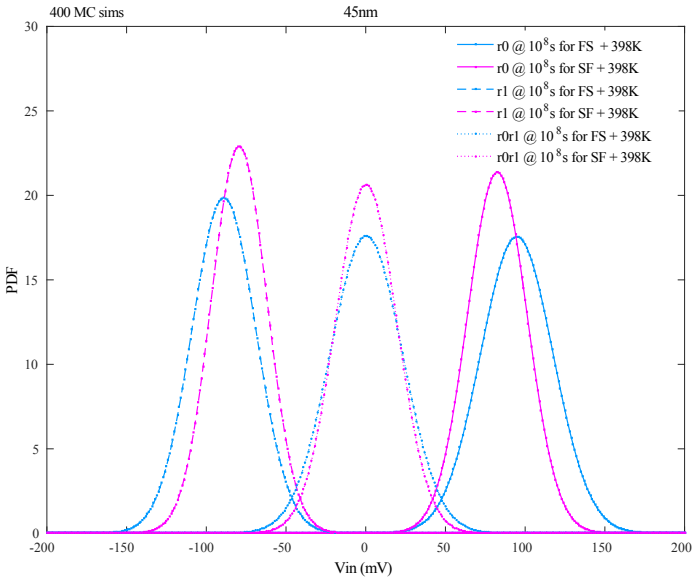


Fig. 8. Temperature impact at time-dependent for both *FS* & *SF* corner at 398K.

(this increases up to 77.5%). Obviously, the offset voltage specification increment is much higher for the *FS* process corner, unbalanced workload, and higher temperature. Note that applying a balanced workload can reduce the impact of offset voltage specification at time-dependent and higher temperature (up to 4×) in the worst case.

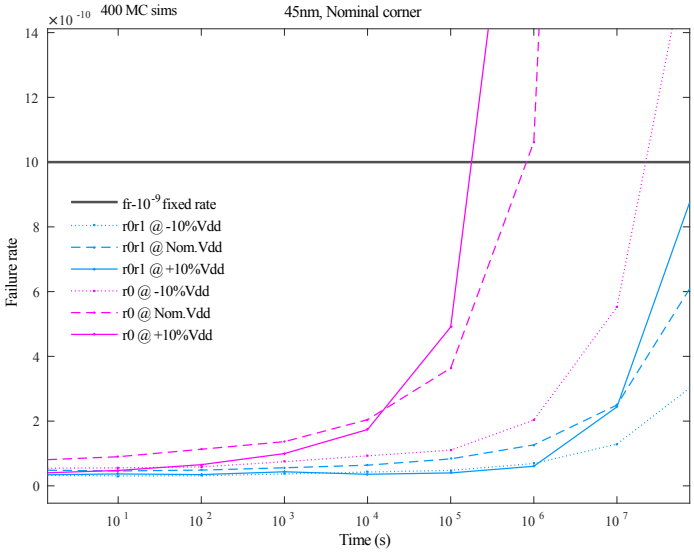


Fig. 9. Voltage impact at time-dependent failure rate analysis for *Nominal* corner.

Fig. 8 depicts the PDF of the offset voltage for extreme temperature conditions, i.e., 398K for both *FS* (worst case) and *SF* (best case) process corners, irrespective of the applied workloads. Fig. 8 shows a similar trend as Fig. 6; however the offset distribution is wider at higher temperature. This is attributed to the severity of temperature impact on the BTI V_{th}. Note that critical temperatures impact the offset voltage more than the supply voltage, irrespective of the process corners and workloads.

7 RELIABILITY FAILURE ANALYSIS

This section performs reliability failure analysis for the sense amplifier and analyzes the failure rate over time. Both the impact of voltage and temperature are considered.

7.1 Methodology

To evaluate the added value of our framework, we will be analyzing the failure rate based on an offset voltage specification which is determined by the previous work. Previously, this offset voltage specification is solely derived from time-zero variations [Agbo et al. 2016]. We add a 10% margin for aging and use our methodology to examine how the failure rate increases over time when actual workloads and aging are considered. The reason for selecting a 10% margin is due to the fact that it is a classical margin for the Design for Manufacturability (DFM) in the ITRS 2005 edition [ITRS 2005; Jeong et al. 2009]. In order to obtain the failure rate f_r over time, Eqn. 6 will be solved for different time steps.

$$f_r(t) = 1 - \int_{V_{in}=-V_{Offset}}^{V_{Offset}} \mathcal{N}(\mu_{MC}, \sigma_{MC}, t) \tag{8}$$

In order to obtain f_r at time t, both V_{Offset} (i.e., the offset voltage specification from time-zero plus 10% margin) and the distribution of the input voltages of the SA $\mathcal{N}(\mu_{MC}, \sigma_{MC}, t)$ must be

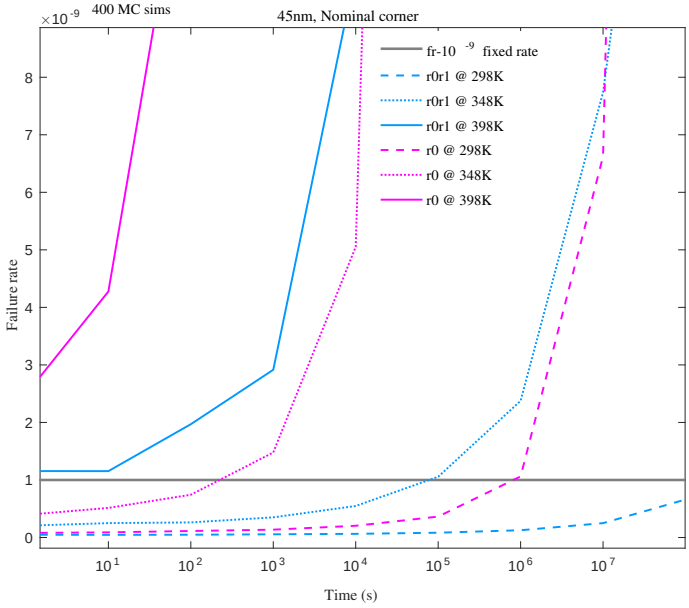


Fig. 10. Temperature impact at time-dependent failure rate analysis for *Nominal* corner.

known. The offset voltage specification at time-zero for a failure rate of 10^{-9} equals 6.1σ using only local process variation [Cosemans 2009]; subsequently, by adding a 10% margin, we obtain the interval values from -V_{Offset} and V_{Offset} of the integral in Eqn 8. For example, in Table 4 the offset specification is 96.1 mV at time 0 (solely based on 6.1σ as average is 0 using a failure rate f_r=10^{-9}) for the *Nominal* corner. Adding the 10% margin leads to an offset specification of 105.7 mV, which is used in the experiments of the following subsections. The distribution of the input offset voltages (i.e., $\mathcal{N}(\mu_{MC}, \sigma_{MC}, t)$) will be simulated for different time steps, similarly as done for 10^8s in Tables 3 and 4. Next, we will analyze the impact of voltage and temperature for different workloads.

7.2 Voltage dependency

Fig. 9 depicts the failure rate for the balanced and unbalanced workloads for the various supply voltages for the *Nominal* process corner. The same 10^{-9} failure rate as for time-zero is used as a reference. Note that for the unbalanced workload only {r0} has been added as {r1} gives similar results. Note that ideally they lead to the same results.

The figure shows that after a certain amount of time, the failure rate starts to increase exponentially. This effect is worse for unbalanced workloads and higher V_{dd}. The 10% added margin to the time-zero offset specification is enough to keep the failure rate below 10^{-9} for balanced workloads. For example, at 10^8s and for -10%V_{dd} the failure rate for the balanced {r0r1} workload is 0.33×10^{-9} which is 67% less than the targeted failure rate, while this is 34% and 3.4% less for Nom.V_{dd} and +10%V_{dd}, respectively. However, for the unbalanced workload at -10%V_{dd}, the failure rate exceeds 10^{-9} by 84.8% after an operational life time of 10^8s, while for the Nom.V_{dd} the failure rate already

increases with 562.5% after 10^7s. Hence, more SAs than expected will fail if previous methods based on time-zero will be used. Therefore, it is important to characterize the degradation accurately and have mitigation schemes to remedy the unreliability of the SA.

7.3 Temperature dependency

Fig. 10 shows a similar graph as Fig. 9 but for different temperatures. The figure also shows that the failure rate strongly increases with the temperature and that the time-zero approach fails significantly. The figure shows that at 398K, also *balanced* workloads do not meet the target of 10^{-9} failure rate after 10^8s. The above results can be even worse for different process corners. Therefore, our presented framework is required to accurately determine the offset voltage specifications.

8 DISCUSSION

Memory sense amplifier robustness and reliability are very important for the overall design of memory systems. The presented analysis showed that the offset voltage specification and failure rate of the SA design are dependent on the process corner, supply voltage, temperature, workload etc. After the evaluation of the simulation results, we conclude the following:

- The offset voltage specification at time-zero variability is marginal for varying supply voltage and this is significant for varying temperature; however, the offset voltage specification is higher for the *FS* corner as compared to the other process corners.
- The offset voltage specification at time-dependent variability and at varying supply voltage for the 5 process corners maintain a unique trend with respect to the average shift, standard deviation and its offset shift; however, the *FS* process corner proves to be the worst corner while *SF* is the least impacted corner.
- The offset specification at time-dependent variability and at varying temperature for the various process corners, irrespective of the workloads follow the same trend; however, the offset specification is more severe at higher temperature and for unbalanced workloads, irrespective of the process corner being considered. *Note that efficient application of balanced workload can reduce the impact of offset specification at time-dependent variability and* 398*K* (up to 4×), irrespective of the process corner considered.
- The proposed technique for offset voltage computation is unique not only in the sense that it utilizes an accurate BTI model and involves the workload dependency for various process corners, but also because it integrates both the zero-time variability as well as time-dependent variability. The extracted results clearly show that using only environmental time-dependent variability for various process corners while considering zero-time variability analysis is not accurate enough. The change in offset voltage between time-zero and time-dependent variability for the different process corners is maximum for *FS* corner and minimum for the *SF*. Hence, the other three process corners fall in between the two extreme corners.
- In addition, the dependency of offset voltage on workload (application) has been depicted to be significant. Applying balanced workload results in reduced impact. Thus, thinking about integrating some features in the circuits to internally create a balance workload during the lifetime of the application is crucial for optimal and reliable designs. Schemes such as bit-flipping [Gebregiorgis et al. 2015] can be useful.
- Our failure analysis results in Section 7 showed that the observed failure rate can be much higher than expected when factors such as workloads, voltage, temperature etc. are not properly considered. Hence, insufficient margins will increase the number of failing devices with needlessly high costs.

- The failure rate results show that adding a 10% margin for the offset voltage specification at time-zero is typically sufficient. However, the failure rate is much worse at high temperatures, irrespective of the workloads and added margins.
- Finally, it is worth noting that the described technique of Fig. 2 can be extended to any digital circuit, as long as there is an obvious metric (for example offset specification) to be quantified. For instance, the critical paths in a pipeline stage can be computed based on the path delay metric.

9 CONCLUSION

This paper investigated an accurate methodology to thoroughly analyse the impact of time-zero and time-dependent variation on offset voltage specification for an SRAM SA design utilizing 45 nm PTM high performance library. The technique considers also the sensitivity analysis, five global process corners (for both time-zero and time-dependent) while taking into account different supply voltages, temperatures, and workloads, and failure analysis as well as degradation as a result of aging. The results of this investigation show that at time zero variation the SA offset voltage specification sensitivity analysis is clearly dominated by the pull down transistors which means that SA designers should be more concerned on the pull down transistors for a better SA offset voltage specification quantification. Furthermore, at time zero variability the offset voltage specification is marginally dependent on varying supply voltages and temperatures. However, this becomes more significant when time-dependent variability for unbalanced workloads are taken into account which implies that balanced workloads can reduce the offset voltage specification impact (up to 4×) in the worst case. In addition, at time-dependent variability the SAs failure rate are mainly not met, hence it is critical to characterize the degradation accurately and provide mitigation techniques to compensate the SA unreliability. The proposed technique gives designers a preferable way of ascertaining the offset voltage specification; hence, aiding designers in making appropriate design choices based on the workloads and environmental conditions.

REFERENCES

M. H. Abu-Rahma, Y. Chen, W. Sy, W. L. Ong, L. Y. Ting, S. S. Yoon, M. Han, and E. Terzioglu. 2011. Characterization of SRAM sense amplifier input offset for yield prediction in 28nm CMOS. In *2011 IEEE Custom Integrated Circuits Conference (CICC)*. Custom Integrated Circuits Conference, Qualcomm Incorporated, San Diego, CA 92121, 1–4. https://doi.org/10.1109/CICC.2011.6055315

I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, and W. Dehaene. 2016. Quantification of Sense Amplifier Offset Voltage Degradation due to Zero-and Run-Time Variability. In *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Mekelweg 4, 2628 CD Delft, The Netherlands, 725–730. https://doi.org/10.1109/ISVLSI.2016.30

M. A. Alam and S. Mahapatra. 2005. A comprehensive model of PMOS NBTI degradation. In *2005 Microelectronics Reliability*, Vol. 45. Elsevier Ltd, USA, 71–81. https://doi.org/10.1016/j.microrel.2004.03.019

S. Borkar. 1999. Design challenges of technology scaling. *IEEE Micro* 19, 4 (Jul 1999), 23–29. https://doi.org/10.1109/40.782564

S. Borkar. 2004. Microarchitecture and Design Challenges for Gigascale Integration. In *Microarchitecture, 2004. MICRO-37 2004. 37th International Symposium on*. IEEE, USA, 3–3. https://doi.org/10.1109/MICRO.2004.24

Cadence. 2016. Spectre Circuit Simulator datasheet. (2016). Retrieved 2016 from http://cadence.com

A. Carlson. 2007. Mechanism of Increase in SRAM V_{min} Due to Negative-Bias Temperature Instability. *IEEE Transactions on Device and Materials Reliability* 7, 3 (Sept 2007), 473–478. https://doi.org/10.1109/TDMR.2007.907409

Mihir Choudhury, Vikas Chandra, Kartik Mohanram, and Robert Aitken. 2010. Analytical Model for TDDB-based Performance Degradation in Combinational Logic. In *Proceedings of the Conference on Design, Automation and Test in Europe (DATE '10)*. European Design and Automation Association, 3001 Leuven, Belgium, Belgium, 423–428.

S. Cosemans. 2009. *Variability-aware design of low power SRAM memories)*. Ph.D. Dissertation. Katholieke Universiteit Leuven, Leuven.

S. Cosemans, W. Dehaene, and F. Catthoor. 2008. A 3.6pJ/access 480MHz, 128Kbit on-Chip SRAM with 850MHz boost mode in 90nm CMOS with tunable sense amplifiers to cope with variability. In *ESSCIRC 2008 - 34th European Solid-State Circuits Conference*. European Solid-State Circuits Conference, 3001 Leuven, Belgium, Belgium, 278–281. https:

//doi.org/10.1109/ESSCIRC.2008.4681846

A. Gebregiorgis, M. Ebrahimi, S. Kiamehr, F. Oboril, S. Hamdioui, and M. B. Tahoori. 2015. Aging mitigation in memory arrays using self-controlled bit-flipping technique. In *The 20th Asia and South Pacific Design Automation Conference*. IEEE, Delft University of Technology, 2628 CD, the Netherlands, 231–236. https://doi.org/10.1109/ASPDAC.2015.7059010

T. Grasser, P. J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel, and B. Kaczer. 2011. Analytic modeling of the bias temperature instability using capture/emission time maps. In *2011 International Electron Devices Meeting*. IEEE, TU Wien, Austria, 27.4.1–27.4.4. https://doi.org/10.1109/IEDM.2011.6131624

S. Hamdioui, D. Gizopoulos, G. Guido, M. Nicolaidis, A. Grasset, and P. Bonnot. 2013. Reliability challenges of real-time systems in forthcoming technology nodes. In *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*. EDA Consortium, San Jose, CA, USA, 129–134. https://doi.org/10.7873/DATE.2013.040

ITRS. 2005. International Technology Roadmap for Semiconductors 2005 [Online]. (2005). Retrieved 2005 from www.itrs.net/common/2005update/2005update.htm

K. Jeong, A. B. Kahng, and K. Samadi. 2009. Impact of Guardband Reduction On Design Outcomes: A Quantitative Approach. *IEEE Transactions on Semiconductor Manufacturing* 22, 4 (Nov 2009), 552–565. https://doi.org/10.1109/TSM.2009.2031789

B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin. 2005. Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification. In *2005 IEEE International Reliability Physics Symposium, 2005. Proceedings. 43rd Annual*. IEEE, B-3001 Leuven, Belgium, 381–387. https://doi.org/10.1109/RELPHY.2005.1493117

B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger. 2010. Origin of NBTI variability in deeply scaled pFETs. In *2010 IEEE International Reliability Physics Symposium*. IEEE, B-3001 Leuven, Belgium, 26–32. https://doi.org/10.1109/IRPS.2010.5488856

B. Kaczer, S. Mahato, V. Valduga de Almeida Camargo, M. Toledano-Luque, P. J. Roussel, T. Grasser, F. Catthoor, P. Dobrovolny, P. Zuber, G. Wirth, and G. Groeseneken. 2011. Atomistic approach to variability of bias-temperature instability in circuit simulations. In *2011 International Reliability Physics Symposium*. IEEE, USA, XT.3.1–XT.3.5. https://doi.org/10.1109/IRPS.2011.5784604

M. Kamal, Q. Xie, M. Pedram, A. Afzali-Kusha, and S. Safari. 2012. An efficient reliability simulation flow for evaluating the hot carrier injection effect in CMOS VLSI circuits. In *2012 IEEE 30th International Conference on Computer Design (ICCD)*. IEEE, USA, 352–357. https://doi.org/10.1109/ICCD.2012.6378663

Seyab Khan, Innocent Agbo, Said Hamdioui, Halil Kukner, Ben Kaczer, Praveen Raghavan, and Francky Catthoor. 2014. Bias Temperature Instability Analysis of FinFET Based SRAM Cells. In *Proceedings of the Conference on Design, Automation & Test in Europe (DATE '14)*. European Design and Automation Association, 3001 Leuven, Belgium, Belgium, Article 31, 6 pages. http://dl.acm.org/citation.cfm?id=2616606.2616644

H. Kukner, S. Khan, P. Weckx, P. Raghavan, S. Hamdioui, B. Kaczer, F. Catthoor, L. Van der Perre, R. Lauwereins, and G. Groeseneken. 2014a. Comparison of Reaction-Diffusion and Atomistic Trap-Based BTI Models for Logic Gates. *IEEE Transactions on Device and Materials Reliability* 14, 1 (March 2014), 182–193. https://doi.org/10.1109/TDMR.2013.2267274

H. Kukner, P. Weckx, J. Franco, M. Toledano-Luque, M. Cho, B. Kaczer, P. Raghavan, Doyoung Jang, K. Miyaguchi, M. G. Bardon, F. Catthoor, L. Van der Perre, R. Lauwereins, and G. Groeseneken. 2014b. Scaling of BTI reliability in presence of time-zero variability. In *2014 IEEE International Reliability Physics Symposium*. IEEE, B-3001 Leuven, Belgium, CA.5.1–CA.5.7. https://doi.org/10.1109/IRPS.2014.6861122

S. V. Kumar, K. H. Kim, and S. S. Sapatnekar. 2006. Impact of NBTI on SRAM read stability and design for reliability. In *7th International Symposium on Quality Electronic Design (ISQED'06)*. IEEE, USA, 6 pp.–218. https://doi.org/10.1109/ISQED.2006.73

E. J. Marinissen, Y. Zorian, M. Konijnenburg, Chih-Tsun Huang, Ping-Hsuan Hsieh, P. Cockburn, J. Delvaux, V. RoÅ¿iÄĞ, Bohan Yang, D. SingelÃĪe, I. Verbauwhede, C. Mayor, R. van Rijsinge, and C. Reyes. 2016. IoT: Source of test challenges. In *2016 21th IEEE European Test Symposium (ETS)*. IEEE, USA, 1–10. https://doi.org/10.1109/ETS.2016.7519331

M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers. 1989. Matching properties of MOS transistors. *IEEE Journal of Solid-State Circuits* 24, 5 (Oct 1989), 1433–1439. https://doi.org/10.1109/JSSC.1989.572629

PTM. 2012. Predictive Technology Model. (2012). Retrieved 2012 from http://ptm.asu.edu/

S. Ray, Y. Jin, and A. Raychowdhury. 2016. The Changing Computing Paradigm With Internet of Things: A Tutorial Introduction. *IEEE Design Test* 33, 2 (April 2016), 76–96. https://doi.org/10.1109/MDAT.2016.2526612

D. Rodopoulos, S. B. Mahato, V. V. de Almeida Camargo, B. Kaczer, F. Catthoor, S. Cosemans, G. Groeseneken, A. Papanikolaou, and D. Soudris. 2011. Time and workload dependent device variability in circuit simulations. In *2011 IEEE International Conference on IC Design Technology*. IEEE, B-3001 Leuven, Belgium, 1–4. https://doi.org/10.1109/ICICDT.2011.5783193

S. S. Sapatnekar. 2011. Overcoming Variations in Nanometer-Scale Technologies. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 1, 1 (March 2011), 5–18. https://doi.org/10.1109/JETCAS.2011.2138250

M. Toledano-Luque, B. Kaczer, P. J. Roussel, T. Grasser, G. I. Wirth, J. Franco, C. Vrancken, N. Horiguchi, and G. Groeseneken. 2011. Response of a single trap to AC negative Bias Temperature stress. In *2011 International Reliability Physics Symposium*.

IEEE, B-3001 Leuven, Belgium, 4A.2.1–4A.2.8. https://doi.org/10.1109/IRPS.2011.5784501

J. Vollrath. 2002. Signal margin analysis for DRAM sense amplifiers. In *Electronic Design, Test and Applications, 2002. Proceedings. The First IEEE International Workshop on*. Electronic Design, Test and Applications (DELTA), Infineon Technologies, 123–127. https://doi.org/10.1109/DELTA.2002.994600

Jiajing Wang, A. Singhee, R. A. Rutenbar, and B. H. Calhoun. 2007. Statistical modeling for the minimum standby supply voltage of a full SRAM array. In *ESSCIRC 2007 - 33rd European Solid-State Circuits Conference*. IEEE, University of Virginia Charlottesville, VA 22904, 400–403. https://doi.org/10.1109/ESSCIRC.2007.4430327

R. Wang, J. Dunkley, T. A. DeMassa, and L. F. Jelsma. 1971. Threshold voltage variations with temperature in MOS transistors. *IEEE Transactions on Electron Devices* 18, 6 (Jun 1971), 386–388. https://doi.org/10.1109/T-ED.1971.17207

P. Weckx, B. Kaczer, M. Toledano-Luque, T. Grasser, P. J. Roussel, H. Kukner, P. Raghavan, F. Catthoor, and G. Groeseneken. 2013. Defect-based methodology for workload-dependent circuit lifetime projections - Application to SRAM. In *2013 IEEE International Reliability Physics Symposium (IRPS)*. IEEE, Katholieke Universiteit Leuven, ESAT-MICAS, Leuven, Belgium, 3A.4.1–3A.4.7. https://doi.org/10.1109/IRPS.2013.6531974

N. H. E. Weste and D. Harris. 2010. *CMOS VLSI Design: A Circuits and Systems Perspective* (4th ed.). Addison-Wesley Publishing Company, USA.

S. Zafar, Y. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, and M. Chudzik. 2006. A Comparative Study of NBTI and PBTI (Charge Trapping) in SiO2/HfO2 Stacks with FUSI, TiN, Re Gates. In *2006 Symposium on VLSI Technology, 2006. Digest of Technical Papers*. IEEE, Yorktown Heights, NY 10598, USA, 23–25. https://doi.org/10.1109/VLSIT.2006.1705198

Comparative BTI Analysis for Various Sense Amplifier Designs

2

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Pieter Weckx[1,2] Stefan Cosemans[1] Praveen Raghavan[1] and Francky Catthoor[1,2]
[1]imec vzw., Kapeldreef 75, B-3001, Leuven, Belgium
[2]Katholieke Universiteit Leuven, ESAT, Belgium
{Pieter.Weckx, stefan.cosemans, Ragha, Francky.catthoor}@imec.be

***Abstract*—With the continuous downscaling of CMOS technologies, ICs become more vulnerable to transistor aging mainly due to Bias Temperature Instability (BTI). This paper presents a comparative study of the BTI impact while considering varying supply voltages and temperatures for three memory sense amplifier (SA) designs: low power (LP), mid power/performance (MP), and high performance (HP). As an evaluation metric, the sensing delay (SD) of the three designs is analyzed for various workloads using 45nm technology. The results show that HP SA degrades faster than MP SA and LP SA irrespective of the workload, supply voltage, and temperature. At nominal supply voltage and temperature, HP degrades up to $1.62\times$ faster than MP, and up to $1.94\times$ faster than LP designs for the worst case workload. In addition, the results show that an increase of 10% in power supply has a marginal impact on the relative degradation. In contrast, the results show that a temperature increment significantly worsens the BTI impact. Finally, the results show that for 16nm technology, BTI impact becomes worse and even causes read failures. This clearly indicates that designing for reliability is not only strongly application dependent, but also technology node dependent. Hence, one has to carefully consider the targeted application, design, and technology node in order to provide appropriate solutions.**

***Index Terms*—BTI, NBTI, PBTI, SRAM sense amplifier**

I. INTRODUCTION

In recent decades, CMOS technology has been sustained with aggressive downscaling that severely impacts the variability and reliability of devices [1–3], originating from manufacturing (i.e., time-zero) and run-time operation, (i.e., from temperature, voltage, aging), respectively. [1]. Bias Temperature Instability (BTI) (i.e., Negative BTI in PMOS transistors and Positive BTI in NMOS transistors) is an aging reliability failure mechanism which affects the performance of MOS transistors by increasing their threshold voltage and reducing their drain current (I_d) over the operational lifetime [4,5]. However, studies of such individual devices or small composites do not allow extrapolation of these effects on larger circuits like Static Random-Access Memories (SRAMs). SRAMs, which are an integral part of any SoC today and occupy a large fraction of today's chips, play a major role in performance, and are critical for system robustness [6]; an SRAM system consists of an array of cells, and peripheral circuitry facilitates the read and write access to the cells; examples are row and column address decoders, control circuits, write drivers, and sense amplifiers. In this work we focus on the SRAM sense amplifier which is a key component that ensures that data is correctly read from the cell array.

Limited work has been published in the area of SA degradation. The authors in [7] analyzed the time zero variability of SRAM sense amplifier systems to determine the offset specification for LP designs. However, they did not take the run-time variability into account. In [8], the authors presented a scheme to determine the signal margins for DRAM SAs based on offset distribution measurements. Also for runtime variability limited work has been published. For example, Menchaca *et al.* [9] analyzed the BTI impact on different SA designs (i.e., current and voltage mode SA) implemented in 32nm technology node using failure probability (i.e., flipping a wrong value) as a reliability metric. Agbo *et al.* [10] investigated the BTI impact on LP drain-input latch type sense amplifier design implemented in 90nm, 65nm, and 45nm for different supply voltages by using sensing delay and sensing voltage as reliability metrics. The same authors [11] investigated the integral impact of BTI and voltage temperature variation on HP standard latch type SA using sensing delay as reliability metric. From the above, we conclude that no prior work discussed the run-time variability of different types of sense amplifiers (i.e., HP, MP, and LP).

This paper focuses on the aging analysis of three different SRAM sense amplifier designs using the realistic atomistic model calibrated with representative test data. The method in [7] is adopted to determine the SA offset specification of all SAs. Subsequently, the BTI impact for each design is analyzed using different workloads, temperatures, voltages, and technology nodes. The results are analyzed and compared relatively to each others.

The rest of the paper is organized as follows. Section II introduces the BTI mechanism, its model, and the architectures of the sense amplifier designs. Section III provides our analysis and simulation framework, and presents also the performed experiments. Section IV analyzes the result for different sense amplifier designs, varying supply voltages, temperatures and technology nodes. Finally, Section V concludes this paper.

II. BACKGROUND

This section briefly discusses bias temperature instability mechanism and its model which is used in this work. Thereafter, it presents the three targeted sense amplifiers.

A. Bias Temperature Instability

The Bias Temperature Instability (BTI) mechanism takes place inside MOS transistors and increases the absolute V_{th}

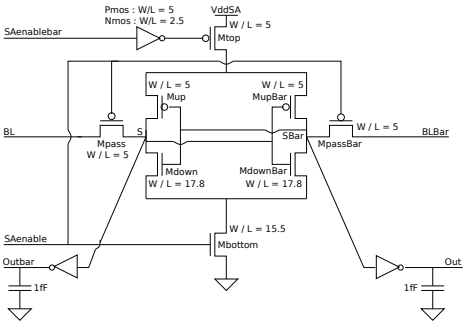


Fig. 1. Standard latch-type Sense Amplifier

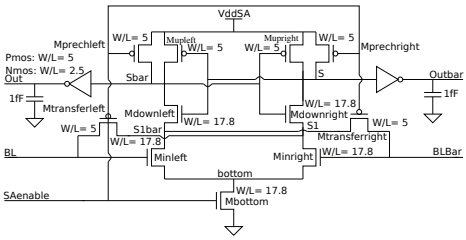


Fig. 2. The Look ahead type Sense Amplifier

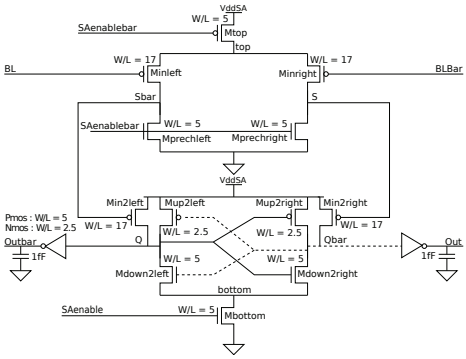


Fig. 3. The Double tail latch type Sense Amplifier

value of the transistors [4,12]. The $V_{\rm th}$ increment in a PMOS transistor occurs under *negative* gate *stress* and is referred to as NBTI, while in an NMOS transistor this occurs under *positive* gate *stress*, and is known as PBTI. Note that for a MOS transistor, there are two BTI phases, i.e., the stress phase and the relaxation phase.

Recently, exhaustive efforts have been put to understand NBTI [4,5,12]. Kaczer *et al.* in [12,14] have analyzed NBTI using an atomistic model. Alam *et al.* [4] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion process. In this work, we use the atomistic model as it produces more accurate results than the RD model [13]. The atomistic model is based on the capture and emission of single traps during stress and relaxation phases of NBTI/PBTI respectively. The threshold voltage shift of the device ΔV_{th} is the accumulated results of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by [13]:

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c+\tau_e}\left\{1-exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{STRESS}\right]\right\} \quad (1)$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c+\tau_e}\left\{1-exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{RELAX}\right]\right\} \quad (2)$$

where τ_c and τ_e are the mean capture and emission time constants, and t_{STRESS} and t_{RELAX} are the stress and relaxation periods, respectively. Furthermore, BTI induced V_{th} is an integral function of Capture Emission Time (CET) map [15], workloads, duty factor and transistor dimensions, which gives the mean number of available traps in each device.

B. Sense Amplifiers

In this paper, we select three sense amplifier designs: standard latch-type sense amplifier is representattive for HP industrial SA designs [7], look-ahead type sense amplifier [16], and double-tail latch-type sense amplifier [17]. They target high performance, inter-mediate performance/power and low power respectively.

Figure 1 depicts the standard latch-type sense amplifier [7], it amplifies a small voltage difference between BL and BLBar during read operations, and produces them at the output (i.e., *Out* and *Outbar*). The positive feedback loop (created by cross-coupled inverters) ensures low amplification time and produces the read value at its output.

Figure 2 depicts the Look-ahead type SA [16] which consists of two stages. In the first stage, when SAenable=0V, S and Sbar are charged up and the inputs BL and BLBar are passed to nodes S1 and S1bar, respectively. In the second stage, the cross-coupled inverters will start the amplification process.

Figure 3 introduces the double-tail latch-type SA [17]. It uses two tails, one for capturing the input and one for amplification and latching. Initially, when SAenable=0V, Mtop and Mbottom are disabled. Thereafter, when SAenable=1V, the capturing tail will charge up nodes S and Sbar; their charge time depends on the inputs BL and BLBar. The ΔS=$V_S - V_{Sbar}$ creates a voltage difference at Q and Qbar through transistors Min2left and Min2right. The amplification and latch tail will amplify this voltage difference.

III. ANALYSIS FRAMEWORK

In this section, the analysis framework, the sense amplifier offset specification and the conducted experiments are presented.

A. Framework Flow

Figure 4 depicts our generic simulation framework to evaluate the BTI impact on the three considered sense

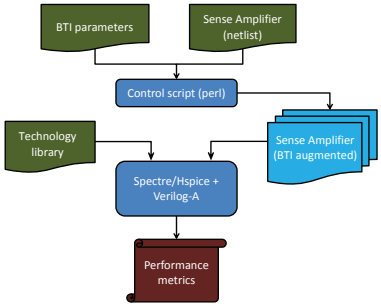


Fig. 4. Analysis framework.

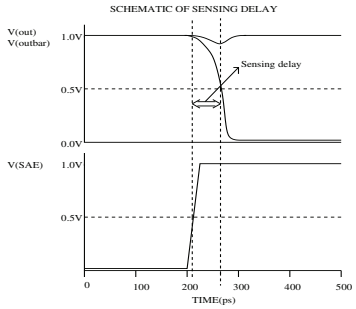


Fig. 5. Metric diagram of Sensing delay.

amplifier designs. It uses HSPICE/Spectre simulator and has the following components.

Input: The general input blocks of the framework are the technology library, Sense Amplifier design, and BTI input parameters. They are explained as follows.

- Technology library: Two technology nodes are considered in this work, they are 16nm and 45nm obtained from PTM library cards [19]. For the LP SA we use the LP library, while for the MP SA and HP SA the HP library. Note that in general any library card can be used.
- SA design: Generally, any sense amplifier design can be used. In this paper we focus only on the standard latch-type SA, look-ahead type SA, and double-tail latch-type SA. The SA designs are described by a SPICE netlist.
- BTI parameters: The BTI induced degradation depends strongly on the stress time duration. The stress time defines how long the workload sequence is being applied. The workload sequence is assumed to be replicated once completed until the age time is reached. To perform realistic workload analysis, we assume that today's application consists of $10\%-90\%$ memory instructions and the percentage of read instructions is typically 50%-90%. Furthermore, we derive from these the two extreme workload cases: worst case: $0.9{*}0.9{=}81\%$ and best case: $0.1{*}0.1{=}1\%$. In addition, we define a mid case workload. This leads to the following workload sequences: $R0^4I^1$ for worst-case, $R0R1I^{198}$ for best-case, and $(R0)I^{24}$ for mid-case, where, $R0$ stands for read 0, $R1$ stands for read 1, I for idle operation (which includes memory write operations). For example, $R0R1I^{198}$ is a workload describing a read 0 and a read 1 performed after each 198 idle operations. For each transistor, we extract its individual duty factor based on the workload and waveform analysis. This enhances the accuracy of our simulation results. This block also specifies the temperatures and voltages.

Processing: Based on the transistor dimensions and other specified inputs, a Control script (perl) generates several instances of BTI augmented SRAM sense amplifier designs. Every generated instance has a distinct number of traps (with their unique timing constants) in each transistor, and are incorporated in a Verilog-A module of the SA netlist [20]. The module responds to the every individual trap, and alters the transistors concerned parameters such as Vth. After inserting BTI in every transistor of the SA design, a Monte Carlo (MC) simulation is performed at different time steps (100 runs at each time step) where circuit simulator (HSPICE/Spectre) is used to investigate the BTI impact.

Output: Finally, statistical post-analysis of the results are performed for varying supply voltages, temperatures and technology nodes in MATLAB environment. The raw outputs are measured directly from HSPICE/Spectre and used to determine the sensing delay. The sensing delay metric is determined when the $trigger$ signal (i.e., sense amplifier enable input signal) reaches 50% of the supply voltage and the $target$ (i.e., either out or outbar falling output signal) reaches 50% of the supply voltage. The difference between the $target$ and the $trigger$ signals is the sensing delay as shown in Fig. 5. Furthermore, the impact of the BTI degradation is measured as the relative difference between the sensing delay with and without BTI.

B. SA offset specification

SA offset voltage is crucial for the correct operation of any SA design. In this work, three SA designs are considered and each of them requires a minimum offset voltage for its operation. In order to accurately determine the SA offset voltage, the impact of process variation on V_{TH} (which is the most critical) is determined, which is thereafter used to measure the impact on the SA offset voltage. The sigma of the V_{TH} distribution is given by:

$$\sigma_{V_{TH0}} = \frac{A_{\Delta V_{TH}}}{\sqrt{2WL}} \tag{3}$$

where $A_{\Delta V_{TH}}$ is the Pelgroms constant [18], W and L the transistor width and length, respectively. While taking this into consideration for each transistor, 100 Monte Carlo simulations are performed on each SA. The minimum offset voltage of a specific SA instance is the voltage difference between SA inputs (Bit lines) where the cross-coupled inverters of the SA remain in their metastable point. This offset voltage is

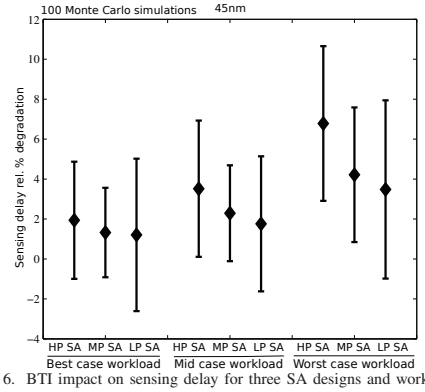


Fig. 6. BTI impact on sensing delay for three SA designs and workloads.

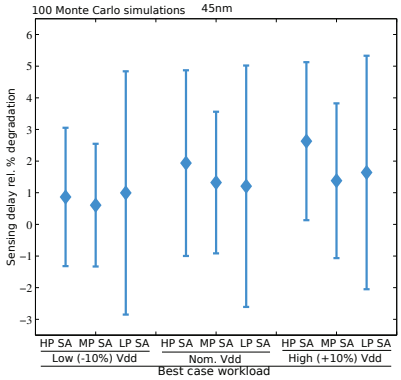


Fig. 7. Supply voltage dependent Sensing delay for best case workload.

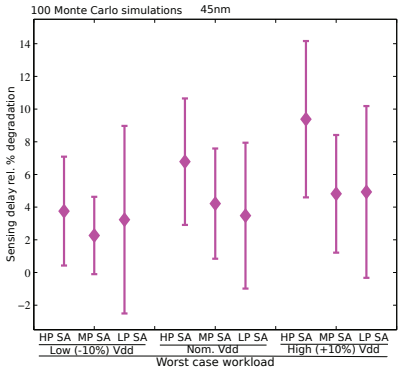


Fig. 8. Supply voltage dependent Sensing delay for worst case workload.

determined by applying a binary search on the input voltage while monitoring the SAs output. This procedure is repeated for all the Monte Carlo instantiations. Because in a good SA design, the offset voltage has nearly normal distribution, the selected SA offset voltage to achieve sufficient yield is 6σ for a 10^{-9} failure rate [7]. For our 45nm designs, we determine the offset specifications to be 89.4mV for HP SA, 104.08mV for MP SA, and 99.5mV for LP SA, while for 16nm design, the offset specification is 225.8mV for HP SA [7].

C. Experiments Performed

In this paper, four sets of experiments are performed.

1 **BTI Impact Experiments:** BTI impact on sensing delay of three SA designs and for different workloads at nominal supply voltage and nominal temperature is investigated.
2 **Supply Voltage Dependent Experiments:** BTI impact on the sensing delay for different SRAM sense amplifiers for varying supply voltages (i.e., from -10% of V_{dd} to V_{dd} and $+10\%$ of V_{dd}) and two workloads (i.e., best case and worst case) is investigated.
3 **Temperature Dependent Experiments:** BTI impact on sensing delay for different SRAM sense amplifiers for three temperatures (i.e., $298K$, $348K$ and $398K$) and two workloads (i.e., best case and worst case) is explored.
4 **Technology Dependent Experiments:** BTI impact on sensing delay for HP SA for two technology nodes (i.e., $16nm$ and $45nm$) using worst case workload is investigated.

IV. EXPERIMENTAL RESULTS

This section, presents the analysis results of the experiments mentioned in the previous section.

A. BTI Impact Experiments

Figure 6 shows the relative increment of the sensing delay w.r.t. the stress time (aging), up to 3 years degradation (10^8s) for the three SA designs and the three workloads (i.e., best case, mid case and worst case). In order to quantify this delay, we simulate the initial BTI-free SA designs and use their sensing delays as references. In this paper, our analysis focus both on the average (i.e., diamond marker) sensing delay and its distribution (i.e., +/-3σ denoted by the edges of the vertical lines in the figure).

The figure shows that the degradation is strongly dependent on the design type and the workload. HP SA degrades on average faster irrespective of workloads than the MP and LP SA designs. For example after an operation of 10^8s for worst case workload, the BTI induced mean degradation is 6.8% for HP SA while 4.2% and 3.5% for MP SA and LP SA designs, respectively. In addition, the degradation is worst for worst case workload as in this case the SA is stressed most; this applies to all designs. The figure also shows that the degradation distribution maintains a distinctive behavior for each SA and workloads. LP SA degradation distribution is wider than the HP SA and MP SA for the worst case workload. For example after an operation of 10^8s for worst case workload, the BTI induced +3σ variation for LP SA is 4.5% while 3.9% and 3.4% for HP SA and MP SA designs, respectively. Besides, the same trend is observed for best case workload while for mid case workload, HP SA and LP SA BTI induced +3σ are approx. 3.4% and 2.4% for MP SA design.

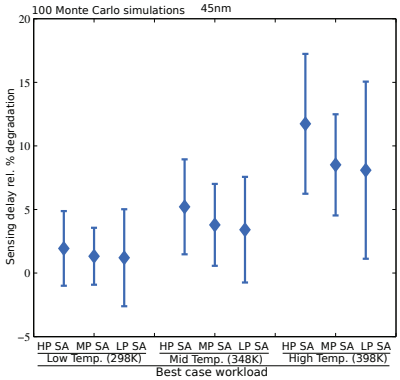


Fig. 9. Temperature dependent Sensing delay for best case workload.

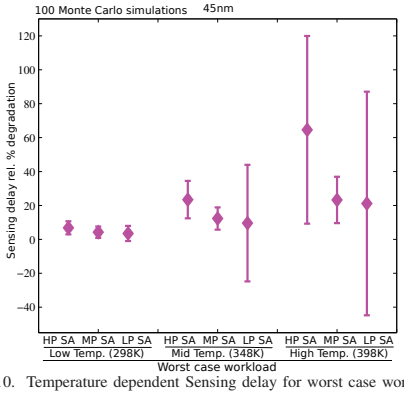


Fig. 10. Temperature dependent Sensing delay for worst case workload.

B. Supply Voltage Dependent experiments

Figures 7 and 8 depict the relative BTI induced sensing delay (i.e., average and +/-3σ variation) for the three SA designs for the various supply voltages and for the best case and worst case workloads, respectively. Note that nominal V_{dd} = 1.0V for HP SA and MP SA, while 1.1V for LP SA.

Figure 7 shows that HP SA is most sensitive to supply voltage; increasing V_{dd} from -10% to +10% leads to a 3× increase in the relative degradation. For example, after an operation of 10^8s for HP SA, the BTI induced mean degradation on sensing delay is 2.63% for +10% V_{dd}, and only 0.87% for -10% V_{dd}. However, the degradation of MP SA and LP SA are marginal. The figure also shows that the degradation distribution maintains a distinctive behavior for each SA and supply voltages. LP SA degradation distribution is the widest irrespective of the supply voltages considered, and the relative delay increment at +3σ equals (w.r.t. to average) 3.8% for 10^8s at nominal V_{dd}.
Figure 8, shows for worst case workload, similar trends as Figure 7 for the relative sensing degradation. However, the degradation is much faster (up to 4.33× faster at -10% V_{dd} to 3.57× at +10% V_{dd} for HP SA, while this is only 3.73× to 3.49× and 3.25× to 3.01× for MP SA and LP SA designs, respectively). In addition the degradation distribution is also wider (up to 1.92× wider at +10% V_{dd} to 1.52× at -10% V_{dd} for HP SA, while this is 1.47× to 1.22× and 1.42× to 1.49× for MP SA and LP SA designs, respectively).

C. Temperature Dependent experiments

Figures 9 and 10 depict the relative BTI induced sensing delay (i.e., average and +/-3σ variation) for the various SA designs and for different temperatures using the best case and worst case workloads, respectively.

Figure 9 shows that the BTI induced relative sensing delay degradation increases with increase in temperature irrespective of the design type. Moreover, HP SA degrades faster (up to 1.4× for MP SA and 1.5× for LP SA designs) at 398K. For example after an operation of 10^8s, the BTI induced mean degradation increases from 1.94% at 298K to 11.74% at 398K for HP SA, while this is only 1.32% to 8.51% and 1.21% to 8.09% for MP SA and LP SA designs, respectively. The figure also shows that the +3σ degradation distribution widens with increase in temperature irrespective of the SA design, and that the LP SA distribution is the widest at all temperatures. For example after an operation of 10^8s at 398K, the BTI induced degradation distribution is 6.96% for LP SA, while this is 5.50% and 3.98% for HP SA and MP SA designs, respectively. Figure 10 representing the results for the worst case workload shows similar trends as Figure 9. However, the degradation is much faster (up to 5.5× at 398K for HP SA). Also the degradation distribution widens (up to 10.06× at 398K for HP SA).

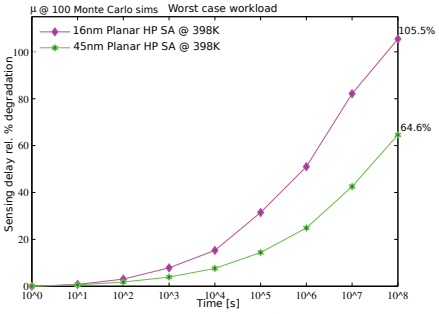


Fig. 11. BTI impact on Sensing delay for two technology nodes.

D. Technology Dependent experiments

Figure 11 depicts the relative BTI induced sensing delay for two technology nodes (i.e., 16nm and 45nm) for HP SA at 398K using worst case workload. Clearly, the BTI induced relative degradation increases significantly at lower nodes, leading to read failures.

Figure 12 depicts the BTI induced read failures for 16nm node for HP SA using worst case workload at 398K; the figure

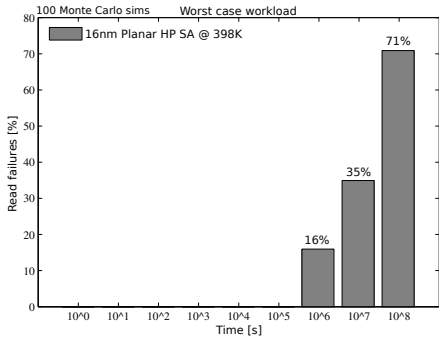


Fig. 12. BTI impact on read failures for 16nm technology node.

also shows the % of read operations that return a wrong value 1 instead of 0. Note that the sensing delay is allowed to be as much as possible. This clearly indicates that using design margin to compensate for degradation is not going to work at lower technology node. It is worth noting that no read failures were observed for 45nm node.

E. Discussion

Memory SA robustness and reliability are very crucial for the overall design of memory systems. The current analysis shows that the degradation and its distribution of each SA design is a function of supply voltage, temperature, workload and technology scaling etc. Examining the simulation results with regard to degradation and its distribution we conclude the following:

- HP SA degrades faster than other SA type regardless of the workload, supply voltage, and temperature; the degradation goes up to 65% (for the 45nm technology considered here); this indicates that BTI may be a serious concern and it may even worst for smaller technologies. Using design margin to compensate for reliability, as it is today, may not be applicable in the future any more.
- The distribution does not maintain the same trend as the degradation with regard to SA types, power, temperature, and workloads. The LP SA distribution widens more than other SA design. This implies a clear trade-off between degradation and distribution in selecting SA design that meet the reliability needs.
- The degradation worsens with scaling for HP SA which causes read failures for the worst case workload; this shows that a more effective mitigation techniques is required at lower technology node.
- The impact of aging on peripheral devices such as SA designs show that there is need to review and analyze the impact of aging on all memory parts and not only the memory cell array as it is usually the case.
- Understanding and quantifying the aging rate of the different parts of memory system is needed for reliable and optimal SRAM designs.
- Designers must be aware of both the average degradation and its distribution for all parts of a memory system (cell array, SA, decoders, write drivers, etc.) and the way they interact together to ensure correct operational life time.
- Designers urgently need to look at mitigation techniques which do take the workload, suppy voltage, temperature, and technology dependence into account.

V. CONCLUSION

This paper investigated the comparative study of the impact of BTI, supply voltage and temperature variation and different workloads for HP SA, MP SA, and LP SA designs for 45nm technology node. In this paper, we have shown that the sensing delay degradation is more impacted by worst case workload for the different SA design at 10^8s operational time. Increasing the supply voltage increases the BTI induced degradation in relative term while reduces in absolute terms leading to more reliable and robust sense amplifiers. Finally, an increase in temperature causes significant increase in sensing delay degradation, thereby severely impacts the reliability of the sense amplifier.

REFERENCES

[1] ITRS, “International Technology Roadmap for Semiconductor 2004” ”www.itrs.net/common/2004 update/2004update.htm.”.
[2] S. Borkar, et al “Micro architecture and Design Challenges for Giga scale Integration”, *Pro. of Intl. Sympos. Micro architecture*, 2004.
[3] S. Hamdioui et al, “Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes”, *DATE*, 2013.
[4] M. A. Alam and S. Mahapatra, “A Comprehensive Model of PMOS NBTI Degradation”, *Microelectronics Reliability* , Vol:45, 2005.
[5] S. Zafar, et al, “A comparative study of NBTI and PBTI in SiO2/HfO2 stacks with FUSI, TiN gates”, *Pro. of VLSI Technology symp.*, 2006.
[6] P. Pouyan, et al, “Process Variability-Aware Proactive Reconfiguration Technique for Mitigating Aging effects in Nano Scale SRAM lifetime”, *IEEE 30th VLSI Test Symposium.*, 2012.
[7] S. Cosemans, “Variability-aware design of low power SRAM memories”, *Ph.D Thesis Katholieke Universiteit Leuven*, 2009.
[8] M. H. Abu-Rahma, et al., ”Characterization of SRAM sense amplifier input offset for yield prediction in 28nm CMOS”, *IEEE CICC*, pp.1-4, Sep. 2011.
[9] R. Menchaca, H. Mahmoodi, “Impact of Transistor Aging Effects on Sense Amplifier Reliability in Nano-Scale CMOS”, *13th Int’l Sym. on Quality Electronic Design*, 2012.
[10] I. Agbo, et al., “BTI Impact on SRAM Sense Amplifier”, *8th IDT*, 16-18 Dec. 2013.
[11] I. Agbo, et al., “Integral Impact of BTI and Voltage Temperature Variation on SRAM Sense Amplifier”, *33rd VTS*, 27-29 Apr. 2015.
[12] B. Kackzar, et al., “Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification”, *IPRS* , pp: 381-387, 2005.
[13] H. Kukner et al, “Comparison of Reaction-Diffusion and Atomistic Trap-based Models for Logic Gates”, *IEEE transactions on device and materials reliability*, 2013.
[14] B. Kaczer, et al., “Origin of NBTI variability in deeply scaled pFETs”, *Reliability Physics Symposium*, 2010.
[15] S. Khan, et al, Bias Temperature Instability Analysis of FinFET based SRAM cells, *DATE*, 2014.
[16] T. Asano, et al, “Low-Power Design Approach of 11FO4 256-Kbyte Embedded SRAM for the Synergistic Processor Element of a Cell Processor”, *IEEE micro*, Oct. 2005, pp. 30-38.
[17] D. Schinkel, et al, “A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time”, *IEEE ISSCC*, 2007, pp. 314-605.
[18] M. J. M. Pelgrom, et al, “Matching properties of MOS transistors”, *IEEE J. Solid-State Circuits* , Vol. 24, no. 5, pp. 1433-1439, Oct. 1989.
[19] Predictive Technology Model ”http://ptm.asu.edu/”,
[20] B. Kaczer, et al., “Atomistic approach to variability of bias-temperature instability in circuit simulations”, *Proc. of IRPS*, April, 2011.

3

BTI Impact and Mitigation on the Memory Read-Path

The content of this chapter incorporates the following research articles:

1. ***I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, Read path degradation analysis in SRAM, 2016 21th IEEE European Test Symposium (ETS), pp. 1–2, May 2016, Amsterdam, The Netherlands.*

2. ***I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, Impact and Mitigation of SRAM Read Path Aging, Microelectronics Reliability Journal, accepted for publication.*

2016 21st IEEE European Test Symposium (ETS)

Read Path Degradation Analysis in SRAM

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Pieter Weckx[1,2] Stefan Cosemans[1] Francky Catthoor[1,2] and Wim Dehaene[2]
[1]imec vzw., Kapeldreef 75, B-3001, Leuven, Belgium
[2]Katholieke Universiteit Leuven, ESAT, Belgium
{Pieter.Weckx, stefan.cosemans, Francky.catthoor}@imec.be
wim.dehaene@esat.kuleuven.be

3

***Abstract*—This paper investigates the impact of aging in the read path of 32nm high performance SRAM; it combines the impact on the memory cell, on the sense amplifier, and on the way they interact. The analysis is done while considering different workloads and by inspecting both the bit-line swing (which reflect the degradation of the cell) and the sensing delay (which reflects the degradation of the sense-amplifier); the voltage swing on the bit lines has a direct impact on the proper functionality of the sense amplifier. The results show that in addition to the sense amplifier degradation, the cell degradation also contributes to the sensing delay increase; the share of this contribution depends on the cell design. Moreover, this sensing delay becomes worst at stressy workloads.**

I. INTRODUCTION

It is well recognized that technology scaling is posing major reliability challenges on electronics reliability [1–3]; e.g., it decreases their lifetime. A common practice in industry is the use of conservative guard-band and application of extra design margins to compensate for the aging impact. Accurate prediction of such impact is crucial for the realization of optimal designs. Obviously, an electronic system consists of different parts; hence, accurate aging prediction needs to consider not only *all* the different parts of the system, but also the way they *interact*, and how they all contribute to the overall degradation of the system. For instance, when it comes to SRAMs (the topic of this paper), predicting the impact of aging by only focusing on the memory array, or by only combining the individual impacts of each components, will lead to optimistic or pessimistic results.

Very limited work is published on the quantification of the degradation impact while considering all the memory components and their interactions. Li [4] analyzed the lifetime prediction of each individual transistor for the *entire* SRAM and for different reliability mechanism (i.e., HCI, TDDB, NBTI). However, this analysis did not involve the workload, which has been shown to have a large impact on the degradation rates [5]. The above clearly shows that an appropriate approach that accurately predict the impact of aging while considering all memory components, the way they interact as well as affect different workloads are required.

This paper analyzes the impact of Bias Temperature Instability (BTI) on the read path consisting of an SRAM cell and sense amplifier (SA). The analysis uses the *Atomistic Model* for aging (which is a calibrated BTI model [6–8]) and considers the *workload dependency* (as the aging variations are strongly workload dependent [5]). To measure both the impact of the cell and SA appropriate workloads are defined while using the bit-line voltage swing and SA sensing delay as metrics. The simulation model used to analyze this impact is explained next.

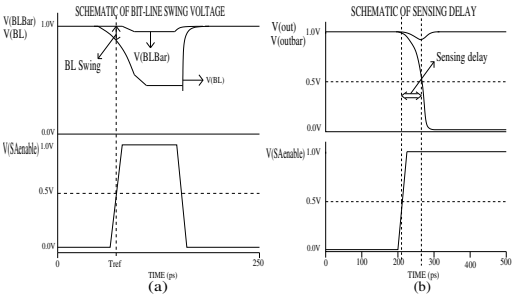


Fig. 1. Metric diagram of (a) Bit-line swing *BLS* and (b) Sensing delay *SD*.

II. SIMULATION MODEL

The model comprises of precharge circuitry, 6T cell, SA precharge, and the SA. These model components *i.e., its netlists*, in addition to the BTI model *Atomistic model* explained in [6–8] are incorporated into the framework flow described in [9,10]. However, the workloads used in the model and its metrics are not described. The workloads and metrics will be explained next.

The workload sequence is assumed to be replicated until the age time is reached. To define the workloads for our analysis, we assume two extreme workloads for the cell's state: (i) 80% zero's where 80% of the cycles the cell holds a zero, and (ii) 20% zero's. Similarly, we assume two workloads for the SA: (i) 80% of the instructions are reads, and (ii) 20% of the instructions are reads. Based on this information, we derive four workload sequences for circuit simulation:

S1: denotes 20% zero's and 80% read instructions for the SA.
S2: i.e., 20% zero's and 20% read instructions for SA.
S3: i.e., 80% zero's and 80% read instructions for SA.
S4: i.e., 80% zero's and 20% read instructions for the SA.

Using the waveform of the read operation and the workload sequences, we extract duty factors for each transistors individually.

To measure the impact of BTI, the following two metrics are used.

978-1-4673-9659-2/16/$31.00 ©2016 IEEE

TABLE I
BTI IMPACT AFTER 10^8S.

Degradation component	Workload	Bit-line swing(mV)	Sensing delay(ps)
Cell-Only	20% zero	107.0	61.09
	80% zero	106.3	61.20
SA-Only	20% read instr.	111.1	61.83
	80% read instr.	111.6	65.71
Combined	S1	107.8	66.08
	S2	107.4	62.18
	S3	107.1	66.21
	S4	106.7	62.29

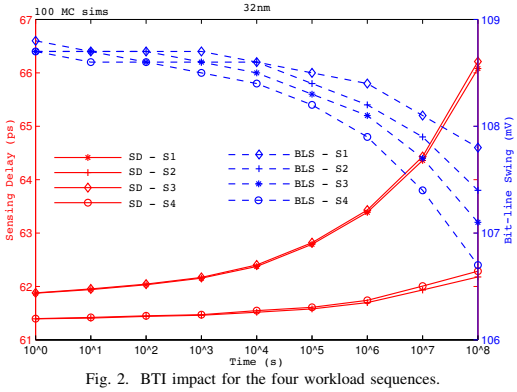


Fig. 2. BTI impact for the four workload sequences.

Bit-line swing: The bit-line swing *BLS* specifies the voltage difference between bit-lines (see Fig. 1a) at a fixed reference time T_{ref}; i.e., the time where the up transition of the sense amplifier enable signal reaches 50% of the supply voltage as shown in Fig. 1a.

Sensing delay: The sensing delay *SD* is the time required for the SA to complete its operation; it is the time between the sense enable activation (i.e., when the up transition reaches 50% of the supply voltage) and the falling output signal of the SA (i.e., when the down transition reaches 50% of V_{dd}) as depicted in Fig. 1b.

III. EXPERIMENTAL RESULTS

Table I shows the results for the three cases for a stress period of 10^8s; the first column presents the simulated case. 'Cell-Only' denotes the case when only the cell is impacted by BTI, 'SA-Only' when only the SA is impacted, and 'Combined' when both the cell and SA degrade due to BTI. Note that in case of 'Cell-Only', both the bit-line swing (*BLS*) and the sensing delay (*SD*) are affected, while in the case of *SA-Only*, the *SD* is impacted (i.e., the *SD* may increase due to slow bit-line swing development or slow SA) while the *BLS* should not be affected. The table reveals the following for the different cases.

For the case Cell-Only, the BLS is marginal dependent on the workload, resulting in almost no impact on the SD. This can be explained by the fact that the pull-down transistors of the cell used for this design are very strong. We will assume SD=61.09ps as the baseline.

For the case SA-Only, the cell is not suffering from BTI; hence, it is not affected and is about 111mV. The SD, however, is affected and increases for more stressy workloads. The SD at 80% read instructions is 6% higher than at 20% reads for which the SD is just 1% more than the baseline.

For the case 'Combined', although the BLS is reduced as compared with the a-fresh cell (see SA-Only case), the dependency of BLS on the workload is marginal due to the chosen design as already mentioned. However, as can be predicted, the results show clear dependency of the SD on the workload; the SD is higher for sequences S1 and S3 which both have 80% read instructions for the SA. At 80% read instructions (S1 and S3), the SD is also 6% higher than at 20% read instructions (S2 and S4); in the latter case the SD is about 2% more than the baseline. Note that the *relative* increase due to workload is the same as for SA-Only' case.

Figure 2 shows how BLS and SD evolve over time for a duration of 3 years degradation (i.e., 10^8s) for the case 'Combined'; each point in the graph corresponds to the average of 100 Monte Carlo simulations. The figure clearly confirms the conclusions extracted from Table I, and that (although in terms of absolute number of our case study the difference are not so big), *the slowest SD* is obtained when *both* the degradation of the cell and the SA are considered. Note that the SD tends to grow very fast when the operational lifetime gets closer to 3 years (10^8s).

IV. CONCLUSIONS

In conclusion, to ensure correct operational lifetime, designers must be aware about how the different parts of the memory degrade, how their interactions contribute to the degradation, and how all of these determine the overall degradation. This will allow for better design margin optimization. Note that in our analysis zero-time variations (process variations) are not taken into account due to the model limitation.

REFERENCES

[1] ITRS, "International Technology Roadmap for Semiconductor 2004" "www.itrs.net/common/2004 update/2004update.htm.".
[2] S. Borkar, et al., "Micro architecture and Design Challenges for Giga scale Integration", *Pro. of Intl. Sympos. Micro architecture*, 2004.
[3] S. Hamdioui et al., "Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes", *DATE*, 2013.
[4] X. Li, et al., "SRAM circuit-failure modeling and reliability simulation with SPICE", *IEEE TDMR*, vol.6, no.2, pp.235-246, Jun. 2006.
[5] P. Weckx, et al., "Defect-based Methodology for Workload-dependent Circuit Lifetime Projections-Application to SRAM", *IRPS*, April 2013.
[6] B. Kaczer, et al., "Atomistic approach to variability of bias-temperature instability in circuit simulations", *Proc. of IRPS*, April, 2011.
[7] B. Kackzar, et al., "Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification", *IPRS* , pp: 381-387, 2005.
[8] H. Kukner et al, "Comparison of Reaction-Diffusion and Atomistic Trap-based Models for Logic Gates", *IEEE transactions on device and materials reliability*, 2013.
[9] I. Agbo, et al., "BTI Analysis of SRAM Write Driver", *10th IEEE International Design & Test Symposium (IDT)*, 14 - 16 Dec. 2015.
[10] I. Agbo, et al., "Comparative BTI Analysis for Various Sense Amplifier Designs", *19th IEEE DDECS*, 20 - 22 Apr. 2016.

Microelectronics Reliability 87 (2018) xxx–xxx

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel

3

Impact and mitigation of SRAM read path aging

Innocent Agbo[a,*], Mottaqiallah Taouil[a], Daniël Kraak[a], Said Hamdioui[a], Pieter Weckx[b,c], Stefan Cosemans[b], Francky Catthoor[b,c], Wim Dehaene[c]

[a] Delft University of Technology, Faculty of Electrical Engineering, Mathematics and CS, Mekelweg 4, Delft 2628 CD, the Netherlands
[b] IMEC vzw, Kapeldreef 75, Leuven 3001, Belgium
[c] Katholieke Universiteit Leuven, ESAT, Belgium

ARTICLE INFO

Keywords:
Bit-line swing
SD
BTI
SRAM sense amplifier

ABSTRACT

This paper proposes an appropriate method to estimate and mitigate the impact of aging on the read path of a high performance SRAM design; it analyzes the impact of the memory cell, and sense amplifier (SA), and their interaction. The method considers different workloads, technology nodes, and inspects both the bit-line swing (*BLS*) (which reflect the degradation of the cell) and the sensing delay (*SD*) (which reflects the degradation of the sense amplifier); the voltage swing on the bit lines has a direct impact on the proper functionality of the sense amplifier. The results with respect to the quantification of the aging, show for the considered SRAM read-path design that the cell degradation is marginal as compared to the sense amplifier, while the *SD* degradation strongly depends on the workload, supply voltage, temperature, and technology nodes (up to 41% degradation). The mitigation schemes, one targeting the cell and one the sense amplifier, confirm the same and show that sense amplifier mitigation (up to 15.2% improvement) is more effective for the SRAM read path than cell mitigation (up to 11.4% improvement).

1. Introduction

CMOS technology scaling is well known for causing crucial reliability challenges on electronics reliability [1-3]; e.g., it reduces their lifetime. A general practice in industry is the use of conventional guard-band and application of extra design margins to counteract for the Bias Temperature Instability (BTI) effect. Accurate estimation of such effect is vital for achieving an optimal design. Clearly, an electronic system comprises of various parts; hence, accurate BTI estimation requires to evaluate not only *all* the various parts of the system, but also the way they *communicate with each other*, and how they all provide to the complete degradation of the system. For example, when it comes to SRAMs, estimating the effect of BTI by only focusing on the memory array, or by only integrating the individual effects of each components, will lead to optimistic or pessimistic results.

Several publications have investigated the impact of reliability on individual SRAM components. Kumar et al. [4] and Carlson et al. [5] analyzed the impact of negative Bias Temperate Stability (NBTI) on the read stability and the Static Noise Margin (SNM) of SRAM cells. Bansal et al. [6] presented insights on the stability of an SRAM cell under the worst-case conditions and analyzed the effect of NBTI and PBTI (positive BTI). Khan et al. [7] performed BTI analysis for FinFET based memory cells for different SRAM designs using SNM, Read Noise Margin (RNM) and Write Triple Point (WTP) as metrics. Menchaca et al. [8] analyzed the BTI impact on different sense amplifier designs implemented on 32 nm technology node by using failure probability (i.e., flipping a wrong value) as a reliability metric. Agbo et al. [9-12] investigated the BTI impact on SRAM drain-input and standard latch-type sense amplifier design, while considering process, supply voltage, and temperature (PVT) variations in the presence of varying workloads and technology nodes. Rodopoulos et al. [13] proposed and investigated the pseudo-transient atomistic-based BTI model with built-in workloads while considering various supply voltages and temperatures. Other research focused on mitigation schemes. For example, Kraak et al. [14] and Pouyan et al. [15] investigated the mitigation of SA offset voltage degradation by considering periodic input switching. Gebregiorgis [16] investigated a low cost self-controlled bit-flipping scheme which reverses all bit positions with respect to an existing bit.

From the above, we conclude that not much work is published on aging, while taking into account all the memory components and thus their interactions, and the effect of mitigation methodologies on the whole memory. Li et al. [17] studied the lifetime estimation of each individual transistor for the *entire* SRAM and for various reliability mechanism (i.e., HCI, TDDB, NBTI). However, this investigation did not require the workload, which has been demonstrated to have a large effect on the degradation rates [13, 18, 19]. In our previous work [20],

* Corresponding author.
E-mail address: i.o.agbo@tudelft.nl (I. Agbo).

https://doi.org/10.1016/j.microrel.2018.05.011
Received 21 June 2017; Received in revised form 16 April 2018; Accepted 16 May 2018

we analyzed the impact of aging in the read path of a 32 nm high performance SRAM design for different workloads. However, the impact of aging on different supply voltages, temperatures, technology nodes, and varying device drive strengths based on *BLS*, *SD*, and energy (E) metrics on the memory read path are yet to be explored. In addition, effective mitigation schemes are not proposed. The above clearly shows that an appropriate approach (that accurately predicts the impact of aging, workloads, and PVT) is needed. Hence, this analysis is crucial to help memory designers understand which of the memory parts to focus on during design for an optimal and reliable design.

In this paper, we set up a step towards this, and we propose an *accurate* method to estimate the impact of Bias Temperature Instability (BTI) on the read path consisting of an SRAM cell and sense amplifier (SA). This enables not only optimal designs (in terms of design margins), but also the development of appropriate design-for-reliability schemes. The proposed method uses the *Atomistic Model* for aging (which is a calibrated BTI model [21, 22]) and considers the *workload dependency* (as the aging variations are strongly workload dependent [18, 19]). To measure both the impact of the cell and SA appropriate workloads are defined while using the bit-line voltage swing, SA *SD*, and energy as metrics. In addition, we analyze different mitigation schemes and their effectiveness.

The rest of the paper is organized as follows. Section 2 provides the SRAM simulation model, and explains BTI mechanism and its model. Section 3 provides the analysis framework and performed experiments. Section 4 analyzes impact of aging on the read path. Section 5 proposes and evaluates the mitigation schemes. Finally, Section 6 and Section 7 discusses the results and concludes this paper, respectively.

2. Background

This section briefly presents the simulation model; it consists of the critical SRAM components in the read path. Finally, it discusses the BTI mechanism and its model.

2.1. Simulation model

Fig. 1 shows the simulation model, which is divided into four parts (i.e., precharge circuitry, 6T cell, SA precharge and the SA). The W/L ratio of each transistor considered for aging is included in the figure. Capacitances are also added to the bit-lines to model the impact of other cells sharing the same column as the simulated cell. Here we assume a 512×128 memory array. During a read operation, first the bit lines are precharged (using precharge circuit), and thereafter one of the bit lines is discharged through one of the cell's pull down transistors of the SRAM 6T cell. The voltage difference/swing is then amplified by the SA to produce the output.

The SA precharge is used to precharge and equalize the data-lines *DL* and *DLBar* to identical voltages before the SA amplifies a small voltage difference between *BL* and *BLBar* during read operations, and produces the output at *Out* (*DL*) and *Outbar* (*DLBar*). The positive feedback loop (created by cross-coupled inverters) ensures low amplification time and produces the read value at its output. Because the considered design is high performance, the cell has *strong pull-down* transistors to speed-up the formation of the swing between the bit-lines during read operation.

It is worth noting that only aging in the cell and the sense amplifier are considered; the cell precharge circuit and the SA precharge circuit are ignored due to their relative large transistor sizes (i.e., less affected by BTI).

2.2. Bias Temperature Instability

The Bias Temperature Instability (BTI) mechanism takes place inside MOS transistors and increases the absolute threshold voltage (V_{th}) value of the transistors [23, 24]. The V_{th} increment in a PMOS transistor occurs under *negative* gate *stress* and is referred to as NBTI, while in an NMOS transistor this occurs under *positive* gate *stress*, and is known as PBTI. Note that for a MOS transistor, there are two BTI phases, i.e., the stress phase and the relaxation phase.

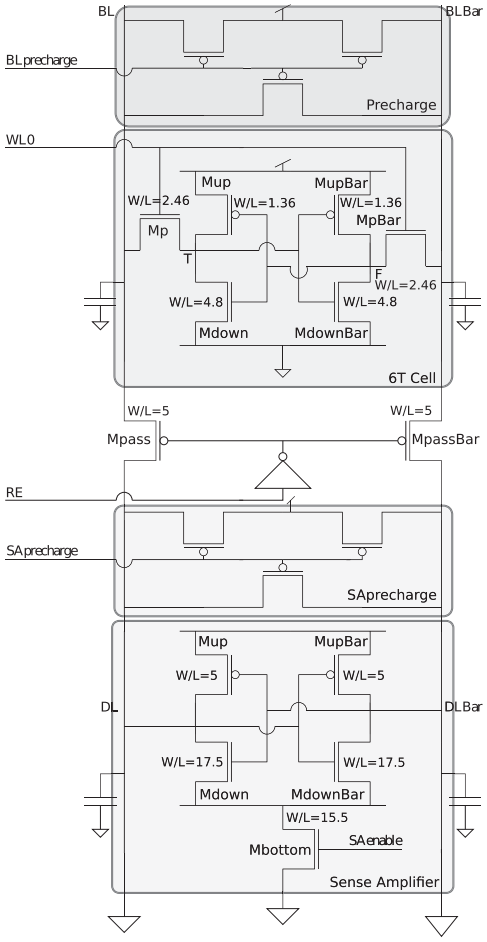


Fig. 1. Simulation setup.

Exhaustive efforts have been put to understand and model BTI appropriately [23-25]. The two most known models are the reaction-diffusion (RD) model proposed by Alam et al. [23], and the atomistic model proposed by Kaczer et al. [21]; the first is deterministic and the second is probabilistic. In this work, we use the atomistic model as it provides more accurate results than the RD model [26]. The atomistic model is based on the capture and emission of single traps during stress and relaxation phases of NBTI/PBTI respectively. The threshold voltage shift ΔV_{th} of the device is the accumulated results of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by [26]:

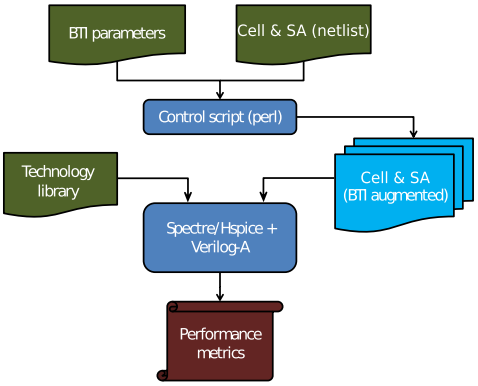


Fig. 2. Analysis framework.

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c + \tau_e}\left\{1 - \exp\left[-\left(\frac{1}{\tau_e} + \frac{1}{\tau_c}\right)t_{STRESS}\right]\right\} \quad (1)$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c + \tau_e}\left\{1 - \exp\left[-\left(\frac{1}{\tau_e} + \frac{1}{\tau_c}\right)t_{RELAX}\right]\right\} \quad (2)$$

where τ_c and τ_e are the mean capture and emission time constants, and t_{STRESS} and t_{RELAX} are the stress and relaxation periods, respectively. Furthermore, BTI induced V_{th} is an integral function of Capture Emission Time (CET) map [7], workloads, duty factor and transistor dimensions, which gives the mean number of available traps in each device, the model also includes the impact of temperature in [21, 22].

3. Analysis framework

This section presents the analysis framework and the conducted experiments.

3.1. Framework flow

Fig. 2 depicts our generic simulation framework to evaluate the BTI impact on the cell and SA designs. It uses Spectre simulator and has the following components.

Input: The general input blocks of the framework are the technology library, cell and sense amplifier design, and BTI input parameters.

- Technology library: In this work we use only the 32 nm PTM library [27]. Note that in general any library card can be used.
- Cell and SA designs: Generally, any memory cell and sense amplifier design can be used. In this paper, we focus only on the design in Fig. 1. The 6T cell and SA designs are described by a SPICE netlist.
- BTI parameters: The BTI induced degradation strongly depends on the stress time duration, hence on the workload. The workload sequence is assumed to be replicated until the age time is reached. To define the workloads for our analysis, we assume two extreme workloads for the cell's state: (i) 80% zero's, that is, 80% of the time the cell holds a zero, and (ii) 20% zero's. Similarly, we assume two workloads for the SA: (i) 80% of the instructions are reads, and (ii) 20% of the instructions are reads. Based on this information, we derive four workload sequences for circuit simulation:
  S1: denotes 20% zero's and 80% read instructions for the SA.
  S2: i.e., 20% zero's and 20% read instructions for SA.
  S3: i.e., 80% zero's and 80% read instructions for SA.
  S4: i.e., 80% zero's and 20% read instructions for the SA.
  Using the waveform of the read operation and the workload sequences, we extract duty factors for each transistors individually.
  It is worth noting that in our investigation time-zero variations (i.e., process variations) are also taken into consideration.

Processing: Based on the inputs (i.e., technology, design, BTI parameters and etc.), a perl control script generates several instances of BTI augmented SRAM cell and/or sense amplifier, depending on the simulation case (see Section 3.2). Every generated instance has a distinct number of traps [21] (with their unique timing constants) in each transistor, and are incorporated in a Verilog-A module of cell netlist only, SA netlist only, or both cell and SA netlists. The module responds to every trap individually, and alters the transistors concerned parameters such as V_{th}. After inserting BTI in every transistor of either coupled design or individual designs, a Monte Carlo (MC) simulation is performed at different time steps (100 runs at each time step) where circuit simulator (Spectre) is used to investigate the BTI impact.

Output: Finally, statistical post-analysis of the results are performed for varying supply voltages, temperatures and device drive strengths in MATLAB environment. The raw outputs are measured directly from Spectre and used to determine the *BLS* and *SD* metrics, which are described next.

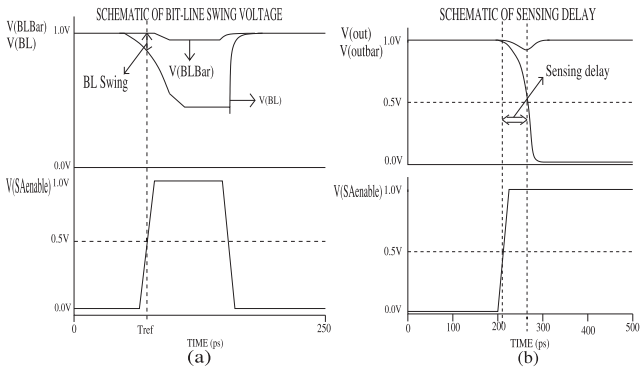


Fig. 3. Metric diagram of (a) *BLS* and (b) *SD*.

Bit-line swing: The *BLS* specifies the voltage difference between bit-lines *BLBar* and *BL* (see Fig. 3a) at a fixed reference time T_{ref}; i.e., the time where the up transition of the sense amplifier enable signal reaches 50% of the supply voltage as shown in Fig. 3a.

Sensing delay: The *SD* is the time required for the SA to complete its operation; it is the time between the sense enable activation (i.e., when the up transition reaches 50% of the supply voltage) and the falling *out* or *outbar* signal (i.e., when the down transition reaches 50% of V_{dd}) as depicted in Fig. 3b.

3.2. Experiments performed

In this paper, four sets of experiments are performed that are related to the quantification of aging, where each set consists of three cases: (a) only the cell degrades (Cell-Only), (b) only the SA degrades (SA-Only), and (c) both of them degrade (Combined).

1. **BTI impact experiments:** BTI impact on *BLS* and *SD* for four workload sequences (*S1*, *S2*, *S3* and *S4*) for 32-nm technology node at nominal supply voltage ($V_{dd} = 0.9V$) and nominal temperature ($T = 298K$) are investigated.
2. **Supply voltage dependent experiments:** BTI impact on the *BLS* and *SD* for varying supply voltages (i.e., from $-10\%V_{dd}$ to $+10\%$ of V_{dd}) and two workload sequences *S2* and *S3* for 32-nm technology node at nominal temperature are investigated. Note that these two sequences present the best and the worst case stresses.
3. **Temperature dependent experiments:** BTI impact on *BLS* and *SD* for three temperatures (i.e., 233*K*, 298*K* and 348*K*) and two workload sequences *S2* and *S3* for 32-nm technology node at nominal supply voltage are explored.
4. **Technology dependent experiments:** BTI impact on *BLS* and *SD* for three technology nodes (i.e., 45-, 32- and 22-nm) and two workload sequences *S2* and *S3* at nominal supply voltages (i.e., $V_{dd} = 1.0\,V$ for 45-nm, $0.9\,V$ for 32-nm, and $0.8\,V$ for 22-nm) and temperature are explored.

4. Experimental results

This section, presents the analysis results of the experiments mentioned in the previous section.

4.1. BTI impact experiments

Table 1 shows the results for the three cases for both time-zero (no aging) and a stress period of 10^8s. Note that the workload is irrelevant for time-zero. The first column presents the simulated case. 'Cell-Only' denotes the case when only the cell is impacted by BTI, 'SA-Only' when only the SA is impacted, and 'Combined' when both the cell and SA degrade due to BTI. Note that in case of 'Cell-Only', both the mean, μ and standard deviation, 3σ of *BLS* and the *SD* are affected, while in the case of *SA-Only*, both the μ and 3σ of the *SD* are impacted (i.e., the μ and 3σ of *SD* may increase due to slow *BLS* development or slow SA) while the *BLS* should not be affected. The table reveals the following for the different cases. For the case 'Cell-Only', the μ BLS is marginally dependent on the workload, resulting in almost no impact on the μ SD while the BLS 3σ increment also impact on the SD's 3σ. This can be explained by the fact that the pull-down transistors of the cell used for this design are very strong (see Fig. 1). We will assume the baseline of SD for μ and 3σ to be equal to 60.80 *ps* and 0.69 *ps*, respectively.

Table 1
BTI impact.

Degradation component	Aging (s)	Workload	Bit-line swing μ (mV)	Bit-line swing 3σ (mV)	Sensing delay μ (ps)	Sensing delay 3σ (ps)
Cell-Only	0	–	108.9	4.2	60.80	0.69
	10^8	20% zero	107.0	6.9	61.09	1.11
	10^8	80% zero	106.3	7.3	61.20	1.18
SA-Only	0	–	110.8	0.7	61.05	1.47
	10^8	20% read instr.	111.1	0.9	61.83	2.88
	10^8	80% read instr.	111.6	1.3	65.71	4.36
Combined	0	–	109.0	4.2	61.24	1.44
	10^8	S1	107.8	6.9	66.08	4.24
	10^8	S2	107.4	6.9	62.18	3.12
	10^8	S3	107.1	7.3	66.21	4.25
	10^8	S4	106.7	7.4	62.29	3.12

For the case 'SA-Only', the cell is not suffering from BTI; hence, it is not affected and is about 111 mV and 0.9 ps for the μ and 3σ of the BLS, respectively. The SD, however, is affected and increases for more stressy workloads. The μ SD at 80% read instructions is ~6% higher than at 20% reads for which the SD is just 1.7% more than the baseline.

For the case 'Combined', although the μ of BLS is reduced as compared with the a-fresh cell (see SA-Only case), the dependency of BLS on the workload is marginal due to the chosen design as already mentioned. However, as can be predicted, the results show clear dependency of the μ of SD on the workload; the SD μ is higher for sequences *S1* and *S3* which both have 80% read instructions for the SA. At 80% read instructions (*S1* and *S3*), the μ SD is also ~6% higher than at 20% read instructions (*S2* and *S4*); in the latter case the μ SD is about ~2.5% more than the baseline. Note that the *relative* increase due to workload is the same as for 'SA-Only' case.

Fig. 4 shows how BLS and SD evolve over time for a duration of 3 years degradation (i.e., 10^8s) for the case 'Combined'; each point in the graph corresponds to the average of 100 Monte Carlo simulations. The figure clearly confirms the conclusions extracted from Table 1, and that (although in terms of absolute number of our case study, the difference are not so big), *the slowest SD* is obtained when *both* the degradation of the cell and the SA are considered. Note that the SD tends to grow very fast when the operational lifetime gets closer to 3 years (10^8s).

4.2. Supply voltage dependency

Table 2 shows the result of Supply Voltage Dependent Experiments for both time-zero (i.e., no aging) and a stress period of 10^8s. The table reveals the following.

For the case 'Cell-Only', similar to the first experiment, the μ and 3σ of the BLS seems to be marginally dependent on the workload. However, a change in the supply voltage clearly influences both the μ of the BLS and the SD and marginally, the 3σ of BLS and SD. Moreover, the

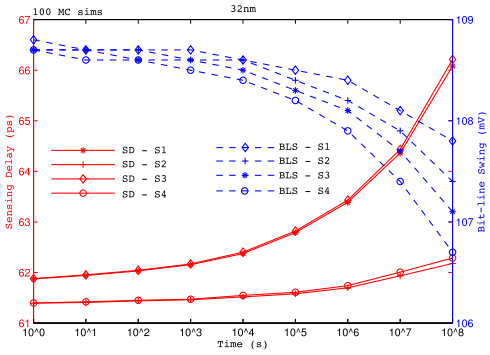


Fig. 4. BTI impact for the four workload sequences.

Table 2
Voltage degradation dependency.

Degradation component	Workload	Aging (s)	Vdd(V)	Bit-line swing μ (mV)	Bit-line swing 3σ (mV)	Sensing delay μ (ps)	Sensing delay 3σ (ps)
Cell-Only	20% zero	0	−10%	77.9	2.7	74.35	0.81
		10^8	−10%	76.7	4.0	74.64	1.02
		0	Nom.	108.9	4.2	60.80	0.69
		10^8	Nom.	107.0	6.9	61.09	1.11
		0	+10%	140.0	6.3	52.12	0.81
		10^8	+10%	136.8	7.9	52.53	1.01
	80% zero	10^8	−10%	76.2	4.5	74.74	1.16
		10^8	Nom.	106.3	7.3	61.20	1.18
		10^8	+10%	135.5	8.0	52.67	1.03
SA-Only	20% read instr.	0	−10%	78.9	0.4	74.74	2.16
		10^8	−10%	79.0	0.6	75.82	3.66
		0	Nom.	110.8	0.7	61.05	1.47
		10^8	Nom.	111.1	0.9	61.83	2.88
		0	+10%	143.4	1.14	52.40	1.50
		10^8	+10%	143.7	1.5	53.18	2.50
	80% read instr.	10^8	−10%	79.3	0.9	80.41	6.28
		10^8	Nom.	111.6	1.3	65.71	4.36
		10^8	+10%	144.4	1.8	58.00	4.84
Combined	S2	0	−10%	78.0	2.79	74.90	2.07
		10^8	−10%	76.9	4.2	75.89	3.70
		0	Nom.	109.0	4.2	61.24	1.44
		10^8	Nom.	107.4	6.9	62.18	3.12
		0	+10%	140.2	6.3	52.79	1.47
		10^8	+10%	137.5	7.6	53.92	2.37
	S3	10^8	−10%	76.7	4.7	80.94	5.52
		10^8	Nom.	107.1	7.3	66.21	4.25
		10^8	+10%	136.8	8.3	58.92	4.63

impact can be higher when time-zero is considered. Increasing the supply voltage accelerates the development of the swing on the bit lines; hence increasing both the μ and 3σ of the BLS. This in turn reduces both the μ and 3σ of the *SD*. On the other hand, reducing the supply voltages reduces the μ BLS, which in turn increases the *SD* μ. A variation of +10% in supply voltage causes an increase of about 26% in the μ BLS and a reduction of about 14% in SD μ, while a variation of −10 % in supply voltage causes a decrease of almost the same percentage in the BLS μ (27%) and an increase of more than 22% in SD μ.

For the case 'SA-Only', although the cell is not suffering from BTI, the supply voltage clearly impacts both the μ and 3σ of the BLS. It follows the same trend as for Cell-Only case. On the other hand, the μ SD is both supply voltage and workload dependent while SD 3σ does not maintain the same trend particularly for stressy supply voltage and workload. A higher voltage improves (reduces) the μ SD, while a lower voltage worsens (increases) the μ SD while the same trend is not followed for the SD 3σ. A +10% variation in V_{dd} causes a reduction of about 13% in μ SD, and −10 % variation in V_{dd} causes an increase of about 22.4% in μ SD. In addition, although the development of voltage swing is accelerated at higher supply voltage, the impact of the workload dependency seems to be slightly higher at higher supply voltage. For example, at −10 % V_{dd} and 10^8s the mean SD increases from 75.82 ps (for 20% read instructions) to 80.41 ps (for 80% read instructions); an increase of 6% while the 3σ SD increases with 71.6%. However, this is about 9% at +10% V_{dd} for the μ SD. Moreover, this is slightly higher by 1.7% at +10% V_{dd} and at time-zero. Note that the impact of supply voltage variation is much dominant than the impact of BTI; this is due to the sizing of the cell's pull-down transistors (see Section Discussion).

For the 'Combined' case, the results show similar trends as for 'SA-Only' case. Even in terms of absolute numbers, the impact of V_{dd} variations and workloads on μ SD are very close (max 1.5% increase) to the results found for 'SA-Only'. Although the slowest SD is obtained in this case, the additional contribution of interaction between degrading cell and degradation SA to the SD as compared with 'SA-Only' is very marginal and does not exceed 1.5%.

Fig. 5 shows how BLS and SD supply voltage dependency evolve over time for a duration of 3 years degradation for the case ' Combined'

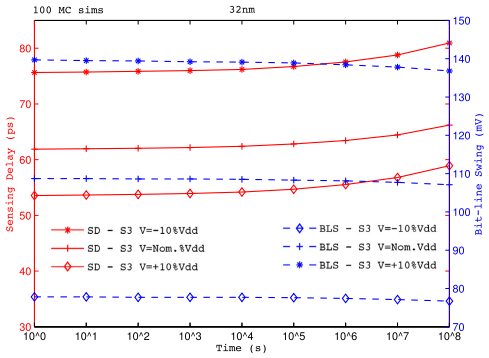


Fig. 5. Supply voltage dependency of SD and BLS for S3 sequence.

using *S3* (worst case stress). The figure shows the impact on the BLS becomes visible when the operational life becomes close to 3 years, which clearly start then impacting the μ SD.

4.3. Temperature dependency

Table 3 shows the results of the Temperature Experiments for both time-zero and a stress period of 10^8s. The table reveals the following.

For the case 'Cell-Only', similar to the first two experiments, at 10^8s the μ and 3σ of BLS seem to be marginally dependent on the workload. However, the temperature strongly influences both the μ and 3σ of BLS and SD. The higher the temperature, the lower the BLS and the higher the SD. Increasing the temperature from 298 K to 348 K reduces the μ and 3σ of BLS with about 33% and 13%, respectively. The increase in temperature also increases the μ and 3σ of *SD* with about 41% and 25%, respectively. Moreover, the reduction in μ and 3σ for BLS will be slightly higher at time-zero while there is a significant increase in *SD*

Table 3
Temperature degradation dependency.

Degradation component	Workload	Aging (s)	Temp. *(K)*	Bit-line swing μ (mV)	Bit-line swing 3σ (mV)	Sensing delay μ (ps)	Sensing delay 3σ (ps)
Cell-Only	20% zero	0	233	176.2	4.8	38.14	0.60
		10^8	233	175.4	6.1	38.23	0.78
		0	298	108.9	4.2	60.80	0.69
		10^8	298	107.0	6.9	61.09	1.11
		0	348	76.2	3.3	85.42	0.69
		10^8	348	72.2	6.1	86.27	1.39
	80% zero	10^8	233	175.2	6.7	38.25	0.85
		10^8	298	106.3	7.3	61.20	1.18
		10^8	348	70.7	6.6	86.62	1.51
SA-Only	20% read instr.	0	233	177.9	1.2	38.19	0.63
		10^8	233	177.9	0.9	38.36	0.79
		0	298	110.8	0.7	61.05	1.47
		10^8	298	111.1	0.9	61.83	2.88
		0	348	77.8	0.5	86.17	3.33
		10^8	348	78.6	1.3	90.10	9.02
	80% read instr.	10^8	233	178.0	1.1	38.63	0.94
		10^8	298	111.6	1.3	65.71	4.36
		10^8	348	79.6	1.9	143.94	113.50
Combined	S2	0	233	176.3	4.8	38.37	0.93
		10^8	233	175.5	6.3	38.59	1.24
		0	298	109.0	4.2	61.24	1.44
		10^8	298	107.4	6.9	62.18	3.12
		0	348	76.4	3.3	86.21	3.00
		10^8	348	73.1	6.5	90.46	8.84
	S3	10^8	233	175.3	6.9	38.87	1.39
		10^8	298	107.1	7.3	66.21	4.25
		10^8	348	72.6	6.8	151.77	117.05

(up to 126% for both μ and 3σ).

For the case 'SA-Only', the temperature clearly impacts both μ and 3σ of the BLS although the cell is not suffering from BTI; hence, the temperature impacts both μ and 3σ of the BLS irrespective of BTI. This impact strengthens the degradation of both μ and 3σ of the *SD* due to the BTI. The *SD* is strongly temperature dependent and the situation becomes worst for stressy workloads. At 20% read instructions, the μ SD increases from 61.83*ps* at 298 K to 90.10*ps* at 348 K; an increase of 45% while the increase for the SD 3σ is too significant (up to 213%!). However, the *SD* μ is 119% and the 3σ is much higher for 80% read instructions! Moreover, at time-zero the impact can be slightly higher for the *SD* μ and much higher for the SD 3σ.

For the 'Combined' case, the results show similar trends as to 'SA-Only' case. Even in terms of absolute numbers, the impact of temperature variations and workloads on both μ and 3σ of the *SD* are close to the results found for 'SA-Only'. Although the slowest μ and 3σ of the SD are obtained in this case, the additional contribution of interaction between degrading cell and degrading SA to the μ *SD* as compared with 'SA-Only' is marginal except for the *S3* at 348 K where this is 5.4%. In addition, the impact can be higher while taking time-zero into account.

Fig. 6 shows how BLS and SD evolve over time for a duration of 3 years degradation for workload *S3* in Combined case. The figure clearly confirms the conclusions extracted from Table 3, and that the degradation of the read paths starts to grow exponentially at high temperatures after a stress time of 10^5s.

4.4. Technology dependency

Table 4 shows the results of Technology Dependent Experiment for both time-zero and a stress period of 10^8s. The table also shows BL-swing in columns 5 and 6. The table reveals the following.

For the case 'Cell-Only', similar to the first experiment, the μ and 3σ of the BLS seem to be marginally dependent on the workload, irrespective of the technology nodes considered. However, as the technology node scales down, the development of swing reduces, irrespective of the workload considered. For example, after an operation of 10^8s and 80% zero, the μ of BL-swing is 1.71% reduction for 45-nm, while 2.45% for 32-nm, and 3.64% for 22-nm technology node.

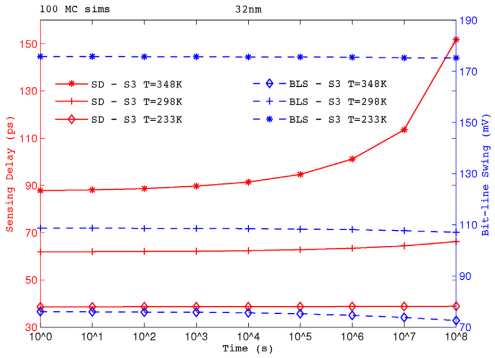


Fig. 6. Temperature degradation dependency of SD and BLS for S3 sequence.

In addition, the 3σ of the BL-swing increases significantly, irrespective of the workloads and technology nodes considered. For example, at 10^8s and 80% zero, 3σ of BL-swing is 30% for 45-nm while 73.81% for 32-nm, and 85.25% for 22-nm. Moreover, the μ of the sensing delay marginally increases in absolute value, while the 3σ of sensing delay relative increment is significant, irrespective of the workloads and technology nodes considered.

For the case 'SA-Only', despite the fact that the cell is not suffering from BTI, the μ and 3σ of BLS increases as well, irrespective of the technology node. On the other hand, as technology node reduces, the μ and 3σ of the sensing delay significantly increases as well. For example, after an operation of 10^8s and for 20% zero, the μ and 3σ of SD is 0.88% and 69.23% for 45-nm, while 1.28% and 95.92%, and 1.94% and 84.42%, for 32-, and 22-nm, respectively. In addition, for 80% zero, the μ and 3σ of SD are 6.74× and 2.73× for 45-nm, while 5.56× and 2.05×, and 5.75× and 2.65×, for 32-, and 22-nm, respectively.

For the case 'Combined', despite the fact that μ of BLS reduces while the 3σ of BLS slightly increases as compared with aged cell (see Cell-

Table 4
Technology degradation dependency.

Tech. nodes	Degradation component	Workload	Aging (s)	BL-swing μ (mV)	BL-swing 3σ (mV)	Sensing delay μ (ps)	Sensing delay 3σ (ps)
45-nm	Cell-Only	–	0	134.9	3.0	66.64	0.62
		20% zero	10^8	133.2	3.8	66.96	0.79
		80% zero	10^8	132.6	3.9	67.06	0.78
	SA-Only	–	0	136.8	1.1	66.80	1.04
		20% read	10^8	137.1	1.3	67.39	1.76
		80% read	10^8	137.6	1.5	70.76	2.75
	Combined	–	0	135.2	3.0	67.13	1.33
		S2	10^8	133.8	3.7	67.95	1.89
		S3	10^8	133.7	4.1	71.44	2.80
32-nm	Cell-Only	–	0	108.9	4.2	60.80	0.69
		20% zero	10^8	107.0	6.9	61.09	1.11
		80% zero	10^8	106.3	7.3	61.20	1.18
	SA-Only	–	0	110.8	0.7	61.05	1.47
		20% read	10^8	111.1	0.9	61.83	2.88
		80% read	10^8	111.6	1.3	65.71	4.36
	Combined	–	0	109.0	4.2	61.24	1.44
		S2	10^8	107.4	6.9	62.18	3.12
		S3	10^8	107.1	7.3	66.21	4.25
22-nm	Cell-Only	–	0	102.5	6.1	57.66	0.98
		20% zero	10^8	99.8	10.3	58.08	1.66
		80% zero	10^8	98.9	11.3	58.22	1.85
	SA-Only	–	0	104.6	1.2	58.20	2.76
		20% read	10^8	104.9	1.9	59.33	5.09
		80% read	10^8	105.4	2.8	64.69	8.93
	Combined	–	0	102.7	6.2	58.47	2.70
		S2	10^8	100.5	10.2	59.81	5.08
		S3	10^8	100.1	11.4	65.66	9.08

Only), the BLS dependency on workload is marginal, irrespective of the considered technology nodes. Nevertheless, the results show a marginal μ dependency of SD on *S2* workload, while a significant μ dependency of SD on *S3* workload, irrespective of the technology nodes considered.

In addition, the results show a significant 3σ dependency of SD on workloads, irrespective of the technology nodes considered. Note that there is marginal relative μ SD increase as compared to 'SA-Only' case while considering various technology nodes. However, the 3σ SD follows the same trend as compared to 'SA-Only' case while considering various technology nodes.

Figs. 7 and 8 depict the relative BTI impact on the SD and the BLS for different technology nodes while considering both worst-case (*S3*) and best-case (*S2*) workloads at nominal supply voltage and temperature. The figures reveal the following:

- The relative sensing delay degradation is more sensitive to technology scaling than BLS. This is also the case for the absolute numbers as the Table 4 shows.
- The impact of workload is more severe on the SD (up to 5×), than the BLS (not more than 1.1×).

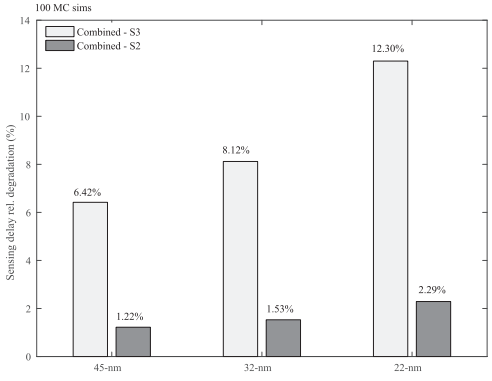


Fig. 7. Technology nodes dependency of SD for both S3 & S2 sequences.

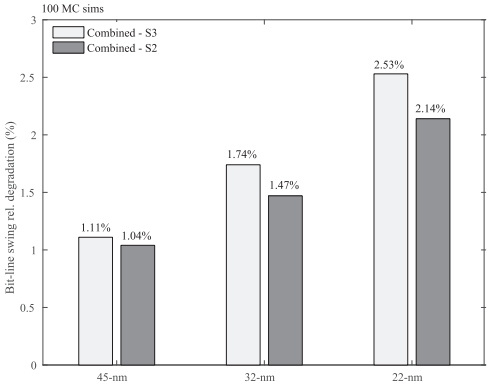


Fig. 8. Technology nodes dependency of BLS for both S3 & S2 sequences.

5. Mitigation schemes

In the previous section, we observed that *BLS* and *SD* may heavily be impacted by BTI. In this section, we investigate two mitigation techniques, i.e., increasing the cell's and the SA's drive strengths. This drive strengths only applies to the pull down transistors for both cell and SA (i.e., *Nom.DS* denoting normal sized transistors, 125%DS denoting 125% larger transistors, and 150%DS denoting 150% larger transistors). Note that the pull up transistors are not affected that much due to bit line and SA pre-charge circuits. We analyze the impact of these drive strengths for workloads *S2* and *S3* which have been defined in Section 3 at nominal supply voltage and temperature conditions. Note that the

Table 5
Cell and SA strength degradation.

Component	Workload	Aging (s)	Device-strength *(DS)*	Bit-line swing μ (mV)	Bit-line swing 3σ (mV)	Sensing delay μ (ps)	Sensing delay 3σ (ps)	Energy μ (fJ)	Energy 3σ (fJ)
Cell-Only	*S2*	0	Nom.	108.9	4.2	61.25	1.59	23.57	0.48
		10^8	Nom.	107.3	6.9	62.22	3.15	23.51	0.81
		0	125%	118.0	4.8	60.06	1.5	23.68	0.42
		10^8	125%	116.7	6.0	61.02	2.37	23.64	0.63
		0	150%	124.8	5.1	58.95	1.65	23.69	0.48
		10^8	150%	123.5	6.9	59.88	2.82	23.65	0.75
	S3	0	Nom.	108.9	4.2	61.27	1.62	23.58	0.48
		10^8	Nom.	107.0	7.2	66.27	4.26	24.36	0.93
		10^8	125%	116.5	6.9	64.68	4.11	24.40	0.93
		10^8	150%	123.2	7.8	63.52	4.32	24.42	1.02
SA-Only	*S2*	0	125%	108.8	4.2	56.86	1.38	22.84	0.42
		10^8	125%	107.2	6.9	57.86	2.40	22.80	0.69
		0	150%	108.7	4.2	54.06	1.29	22.32	0.39
		10^8	150%	107.1	6.9	54.97	2.07	22.28	0.54
	S3	10^8	125%	106.9	7.2	61.82	4.14	23.68	1.02
		10^8	150%	106.8	7.2	58.73	3.54	23.15	0.90
Combined	*S2*	0	125%	117.9	4.8	55.88	1.47	22.98	0.42
		10^8	125%	116.6	6.0	56.74	2.31	22.94	0.69
		0	150%	124.5	5.1	52.11	1.26	22.46	0.39
		10^8	150%	123.2	6.9	52.83	2.04	22.39	0.60
	S3	10^8	125%	116.4	6.9	60.16	3.78	23.68	0.90
		10^8	150%	123.0	7.8	56.23	3.06	23.17	0.78

cell strength influences the BLS and thus indirectly the *SD*. It is worth noting that increasing the device size could lead to an increase in the bit-line length and therefore, also an increase in the delay. Hence, the impact of sizing on the overall latency of the array should be also explored.

Table 5 shows the individual impact of the drive strength of the Cell, the SA and their combined impact for both time-zero and a stress period of 10^8s. In the table, 'Cell-Only' *denotes* the case where only the cell's pull-down transistors drive strength are sized up (i.e., *Nom.*DS, 125%DS and 150%DS). Similarly, 'SA-Only' presents the case where only the drive strength of the pull-down transistors of the SA are sized up. In the 'Combined' case, the pull down transistors of both the cell and SA are simultaneously re-sized. The second column specifies the applied workload, both the cell and SA are stressed using either workload *S2* or *S3*. This workload is applied whether or not a component is re-sized or not. The third column presents the aging (lifetime) while the fourth column specifies the device strength (DS) of the pull down transistors, and the last 6 columns show the results; the evaluated metrics are *BLS*, *SD*, and Energy (E), respectively. The BLS and SD are defined in Section 3, while the energy is defined as the dynamic energy consumption for a single read operation. Next, the three cases will be described.

Cell-Only: For the case 'Cell-Only', the μ BLS significantly increases when the transistors are re-sized while the 3σ remains the same. For example, from 107 mV to 123 mV when a 150% bigger size is used. This 15% μ BLS increment is more or less workload independent. However, the μ BLS increment leads to a much smaller *SD* (both μ and 3σ) improvement. For example, for *S2* this improvement is only $\frac{62.22-59.88}{62.22}\times 100 = 3.7\%$, while $\frac{66.27-63.52}{66.27}\times 100 = 4.1\%$ for workload *S3* for the μ *SD*. Moreover, the 3σ *SD* follows the same trend as the μ *SD*. The energy consumption does not alter much with resizing. Although the operation is faster, the peak power consumption increases while there is a marginal reduction in leakage power.

SA-Only: In contrast, 'SA-Only' has the opposite effect and there is no impact on the *BLS*. However, a higher reduction for both μ and 3σ of the *SD* is observed as compared to the 'Cell-Only'. This delay depends strongly on the applied workload. Furthermore, the device drive strength marginally impacts the energy consumption. For example, at 10^8s increasing the device drive strength from 0% to 150%, has no impact on the μ BLS (small differences are due to Monte Carlo simulations) up to 0.2% while no impact for the 3σ BLS; and marginally reduces the energy consumption up to 5.0%, while μ *SD* significantly reduces with up to 11.4% for the worst-case (*S3*) workload.

Combined: For the 'Combined' case, the results show that both μ and 3σ of the BLS is following the same trend as the 'Cell-Only' and the μ SD only slightly improves with respect to the case 'SA-Only'. For example, the impact difference for a 150% device drive strength (DS) on BLS between 'Cell-Only' and 'Combined' is 0.3 mV, this difference can be attributed to Monte Carlo variations. With respect to the μ *SD*, in the case 'SA-Only' a 150% drive strength is able to achieve a reduction of 11.4%, while this is 15.2% for the combined case. Moreover, at time-zero, there is a slight difference. In addition, the energy consumption is similar as well.

Fig. 9 shows the impact of different device drive strength on both *BLS* and *SD* for the 'Combined' case, for workload *S3*. The figure shows that the BLS marginally reduces over time (i.e., up to 1.47% for Nom.DS, 0.94% for 125%DS and 0.97% for 150%DS) while the SD significantly increases (i.e., up to 7.01% for Nom.DS, 6.82% for 125% DS, and 6.73% for 150%DS) over the operational life time. The relative differences between the different drive strengths are marginal.

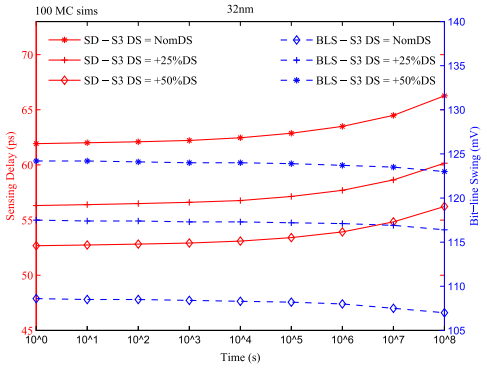


Fig. 9. Cell and SA strength degradation dependency of SD and BLS for S3 sequence.

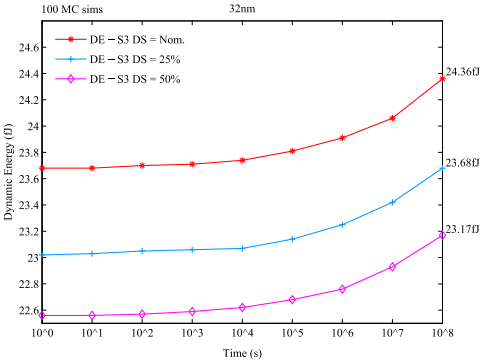


Fig. 10. Cell and SA strength degradation dependency of Dynamic Energy for S3 sequence.

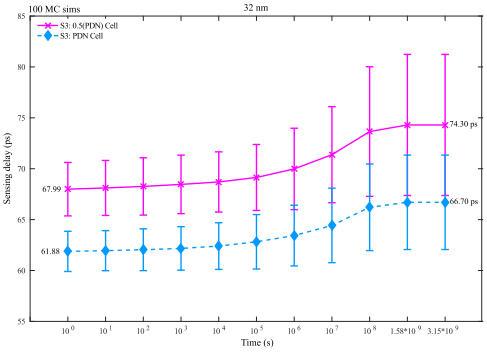


Fig. 11. Variation in SD for S3 sequence for two PDN ratios.

Fig. 10 shows the impact of the device drive strengths on the energy consumption for the 'Combined' case; the energy reduces as the drive strength increases, irrespective of the operational life time. However, the decrease does not exceed 5.0%. For example, at 10^8s and for *DS = Nom*, the energy consumption is 24.36 fJ, while this is 23.17 fJ for *DS = 150%*. In addition, the figure shows for a given drive strength that the aging causes the energy to slightly increase up to 3.0%, irrespective of the drive strength.

Overall, the most effective mitigation technique would be to re-size the SA Only, especially, when the area is also considered. Increasing the cell sizes affects the whole memory matrix, while increasing the 'SA-Only' has a much lower area impact.

6. Discussion

The memory cell and SA robustness are vital for the overall design of memory systems. Below some interesting observations are made.

The obtained results clearly show that for the considered SRAM design the cell has a low impact and that the SA is the major component responsible for the read path timing degradation, even under different voltages, temperatures and technology nodes. Therefore, this information can be used by the designers to optimize the design margins of the cell. One possible explanation of the marginal contribution of the cell degradation to the SD is the cell's strong pull down transistors. Therefore, we investigate the impact of a small cell where we assume W/L of the pull-down transistors to be 2.4 instead of 4.8 (see Fig. 1). The simulation is performed for 10 years using *S3* workload (Combined case), and the results both for the initial design and the smaller cell design (0.5PDN) are shown in Fig. 11. Although the trends of the SD increase for the two simulations seem similar, there are three interesting points to make. First, the relative increase of SD is 7% for the initial design, while this is 9% for the smaller one. Hence, the stronger the pull-down transistors of the cell, the smaller the contribution of the cell to the SD. Second, as the figure shows, the size of the pull-down transistors have also an impact on the SD spread; the stronger the devices, the smaller the spread (i.e., $+/-3\ \sigma$ represented by the boundaries of the vertical lines in Fig. 11). Third, the SD increases relatively faster after 10^4s, but then tends to saturate after 3 years (10^8s); the relative increase from 3 years to 10 years is no more than 0.7% for initial design and 0.9% for smaller version. Clearly the size of the cell's pull-down devices can be used also to minimize the degradation of the read path in SRAMS; and obviously this should be done while considering the SNM of the cell to ensure the stability of the cell as well.

Clearly, reducing the pull-down network (PDN) ratio (e.g., *pull-down transistors*) will only slightly increase the SD (2.0% difference). Hence, the memory cell area can be optimized as long as the SD is within acceptable limit. However, it is crucial to ensure the cell stability for the smaller cell. Therefore, we investigate for both the nominal and the smaller cell three metrics: *HSNM* (hold static noise margin), *RSNM* (read static noise margin), and *WTP* (write trip point) while considering two workloads (i.e., worst case (*WC*) and best case (*BC*)) for 3 years lifetime as shown in Table 6. The *HSNM* is the voltage V_n that flips the cell when it is injected at its internal node; it is swept from $-V_{dd}$ to V_{dd} while the word lines are disconnected from the bit lines. The *RSNM* is the V_n that flips the cell while the word lines are connected to the bit lines and V_n is swept from $-V_{dd}$ to V_{dd}. The *WTP* is the bit line voltage at which the cell flips while the word lines are connected to the bit lines; this voltage can be found by sweeping one of the bit lines potential from $-V_{dd}$ to V_{dd} [7, 28]. Table 6 shows that for both cells *HSNM* marginally reduces after 3 years (does not exceed 3.9%), and that the relative difference is not more than 1.40%, irrespective of the workload and cell size considered. However, the results show that the *RSNM* reduces quite significantly for both cells; this is up to 9.4% and 5.3% for the *WC* and *BC* workloads respectively, irrespective of the cell size. The difference between both cells is marginal. The table finally shows that the *WTP* increases marginally, irrespective of the workload and cell size considered, and that the relative difference between the two cells does not exceed 1.44%. It is worth noting that for the performed experiments, halving the cell size *does not impact the cell stability much as compared to the normal cell size.*

Our next observation is with respect to the impact of supply voltage. Higher voltage increases the bit line swing after an operation of 10^8s and *reduces* the SD. Hence, it can be used to compensate for the degradation of read path especially when the targeted application poses a worst stress on the read path. Obviously, this comes at additional power

Table 6
Cell stability analysis.

	Time (s)	*Nominal cell size*		*Halve PDN ratio*	
		WC	*BC*	*WC*	*BC*
HSNM (mV)	0	312.8	312.8	309.2	309.2
	10^8	300.6	308.3	298.9	304.0
	Rel. %	−3.90	−1.44	−3.33	−1.68
RSNM (mV)	0	168.3	168.3	167.1	167.1
	10^8	152.5	160.0	153.2	158.2
	Rel. %	−9.39	−4.93	−8.32	−5.33
WTP (mV)	0	269.7	269.7	272.6	272.6
	10^8	271.2	277.9	275.1	279.5
	Rel. %	0.56	3.04	0.92	2.53

consumption.

Furthermore, we observed that a higher temperature does not only reduce the BLS (which may impact the functionality) but also significantly increases the SD. Hence, using appropriate cooling is crucial for lifetime extension and degradation retardation.

Moreover, we observed that the degradation is more significant for SD and S3 workload at a lower technology node (*22-nm*); this leads to read failures even at nominal supply voltage. Hence, this implies that there must be a tradeoff between performance and reliability.

Finally, we observed that resizing the cell only marginally mitigates the read path degradation. In contrast, resizing the SA is much more effective. Therefore, more research should focus on effective mitigation schemes for SA, such as input switching in [14].

7. Conclusion

This paper investigated an accurate technique to estimate and mitigate the impact of Bias Temperature Instability (BTI) on the read path of a memory design while considering various degrading components i.e., *Cell only*, *SA only*, and *Combined* (i.e., cell and SA), and for different workloads, supply voltages, temperatures and technology nodes. Hence, the proposed methodology for the entire read path degradation analysis is an interesting case study as it allows for a better understanding of the overall degradation and hence for better design margin optimization. To ensure correct operational lifetime, designers must be aware about how the different parts of the memory degrade, how their interactions contribute to the degradation, and how all of these determine the overall degradation.

References

[1] ITRS, International Technology Roadmap for Semiconductor, 2005 www.itrs.net/common/2005update/2005update.htm, SIA.

[2] S. Borkar, Microarchitecture and design challenges for gigascale integration, MICRO 37 (2004), http://dx.doi.org/10.1109/MICRO.2004.24 3–3.

[3] S. Hamdioui, D. Gizopoulos, G. Guido, M. Nicolaidis, A. Grasset, P. Bonnot, Reliability challenges of real-time systems in forthcoming technology nodes, DATE (2013) 129–134, http://dx.doi.org/10.7873/DATE.2013.040.

[4] S.V. Kumar, C.H. Kim, S.S. Sapatnekar, Impact of NBTI on SRAM read stability and design for reliability, the 7th Int. Symp. Qual. Electron. Des. 0 (9) (2006) 210–218, http://dx.doi.org/10.1109/ISQED.2006.73.

[5] A. Carlson, Mechanism of increase in SRAM VMIN due to negative-bias temperature instability, IEEE TDMR 7 (3) (2007) 473–478, http://dx.doi.org/10.1109/TDMR.2007.907409.

[6] A. Bansal, R. Rao, J.-J. Kim, S. Zafar, J.H. Stathis, C.-T. Chuang, Inpacts of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability, J. Microelectron. Reliab. 49 (6) (2009) 642–649, http://dx.doi.org/10.1016/j.microrel.2009.03.016.

[7] S. Khan, I. Agbo, S. Hamdioui, H. Kukner, B. Kaczer, P. Raghavan, F. Catthoor, Bias Temperature Instability analysis of FinFET based SRAM cells, Des. Autom. Test Eur. Conf. Exhib. (DATE) (2014) 1–6, http://dx.doi.org/10.7873/DATE.2014.044.

[8] R. Menchaca, H. Mahmoodi, Impact of transistor aging effects on sense amplifier reliability in nano-scale CMOS, 13th Interna- tional Symposium on Quality Electronic Design (ISQED), 2012, pp. 342–346, , http://dx.doi.org/10.1109/ISQED.2012.6187515.

[9] I. Agbo, M. Taouil, S. Hamdioui, H. Kukner, P. Weckx, P. Raghavan, F. Catthoor, Integral impact of BTI and voltage temperature variation on SRAM sense amplifier, IEEE 33rd VLSI Test Symposium (VTS), 2015, pp. 1–6, , http://dx.doi.org/10.1109/VTS.2015.7116291.

[10] I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, Comparative BTI analysis for various sense amplifier designs, IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS), 2016, pp. 1–6, , http://dx.doi.org/10.1109/DDECS.2016.7482438.

[11] I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, Quantification of sense amplifier offset voltage degradation due to zero-and run-time variability, IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2016, pp. 725–730, , http://dx.doi.org/10.1109/ISVLSI.2016.30.

[12] I. Agbo, M. Taouil, D. Kraak, S. Hamdioui, H. Kukner, P. Weckx, P. Raghavan, F. Catthoor, Integral impact of BTI, PVT variation, and workload on SRAM sense amplifier, IEEE Trans. Very Large Scale Integr. VLSI Syst. 25 (4) (2017) 1444–1454, http://dx.doi.org/10.1109/TVLSI.2016.2643618.

[13] D. Rodopoulos, P. Weckx, M. Noltsis, F. Catthoor, D. Soudris, Atomistic pseudo-transient BTI simulation with inherent workload memory, IEEE Trans. Device Mater. Reliab. 14 (2) (2014) 704–714, http://dx.doi.org/10.1109/TDMR.2014.2314356.

[14] D. Kraak, I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, Mitigation of sense amplifier degradation using input switching, Des. Autom. Test Eur. Conf. Exhib. (DATE) 2017 (2017) 858–863, http://dx.doi.org/10.23919/DATE.2017.7927107.

[15] P. Pouyan, E. Amat, A. Rubio, Process variability-aware proactive reconfiguration technique for mitigating aging effects in nano scale SRAM lifetime, IEEE 30th VLSI Test Symposium (VTS). 2012, pp. 240–245, , http://dx.doi.org/10.1109/VTS.2012.6231060.

[16] A. Gebregiorgis, M. Ebrahimi, S. Kiamehr, F. Oboril, S. Hamdioui, M.B. Tahoori, Aging mitigation in memory arrays using self-controlled bit-flipping technique, The 20th Asia and South Pacific Design Automation Conference, 2015, pp. 231–236, , http://dx.doi.org/10.1109/ASPDAC.2015.7059010.

[17] X. Li, J. Qin, B. Huang, X. Zhang, J.B. Bernstein, SRAM circuit-failure modeling and reliability simulation with SPICE, IEEE Trans. Device Mater. Reliab. 6 (2) (2006) 235–246, http://dx.doi.org/10.1109/TDMR.2006.876568.

[18] P. Weckx, B. Kaczer, M. Toledano-Luque, T. Grasser, P.J. Roussel, H. Kukner, P. Raghavan, F. Catthoor, G. Groeseneken, Defect-based methodology for workload-dependent circuit lifetime projections - application to SRAM, Int. Reliab. Phys. Symp. (IRPS) (2013), http://dx.doi.org/10.1109/IRPS.2013.6531974 3A.4.1–3A.4.7.

[19] D. Rodopoulos, S.B. Mahato, V.V. de Almeida Camargo, B. Kaczer, F. Catthoor, S. Cosemans, G. Groeseneken, A. Papanikolaou, D. Soudris, Time and workload dependent device variability in circuit simulations, IEEE International Conference on IC Design Technology, 2011, pp. 1–4, , http://dx.doi.org/10.1109/ICICDT.2011.5783193.

[20] I. Agbo, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, Read path degradation analysis in SRAM, 21th IEEE European Test Symposium (ETS), 2016, pp. 1–2, , http://dx.doi.org/10.1109/ETS.2016.7519325.

[21] B. Kaczer, S. Mahato, V.V. de Almeida Camargo, M. Toledano-Luque, P.J. Roussel, T. Grasser, F. Catthoor, P. Dobrovolny, P. Zuber, G. Wirth, G. Groeseneken, Atomistic approach to variability of bias-temperature instability in circuit simulations, Int. Reliab. Phys. Symp. (2011), http://dx.doi.org/10.1109/IRPS.2011.5784604 XT.3.1–XT.3.5.

[22] T. Grasser, P.J. Wagner, H. Reisinger, T. Aichinger, G. Pobegen, M. Nelhiebel, B. Kaczer, Analytic modeling of the bias temperature instability using capture/emission time maps, International Electron Devices Meeting, 2011, http://dx.doi.org/10.1109/IEDM.2011.6131624 27.4.1–27.4.4.

[23] M.A. Alam, S. Mahapatra, A comprehensive model of PMOS NBTI degradation, Physica 45 (1) (2004) 71–81, http://dx.doi.org/10.1016/j.microrel.2004.03.019.

[24] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, M. Goodwin, Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification, IEEE IRPS (2005) 381–387, http://dx.doi.org/10.1109/RELPHY.2005.1493117.

[25] S. Zafar, Y. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, M. Chudzik, A comparative study of NBTI and PBTI (charge trapping) in SiO2/HfO2 Stacks with FUSI, TiN, Re Gates, Symposium on VLSI Technology, Digest of Technical Papers, 2006, pp. 23–25, , http://dx.doi.org/10.1109/VLSIT.2006.1705198.

[26] H. Kukner, S. Khan, P. Weckx, P. Raghavan, S. Hamdioui, B. Kaczer, F. Catthoor, L.V. der Perre, R. Lauwereins, G. Groeseneken, Comparison of reaction-diffusion and atomistic trap-based BTI models for logic gates, IEEE Trans. Device Mater. Reliab. 14 (1) (2014) 182–193, http://dx.doi.org/10.1109/TDMR.2013.2267274.

[27] PTM, Predictive Technology Model, (2008) Beckley, Arizona http://ptm.asu.edu/.

[28] E. Seevinck, F.J. List, J. Lohstroh, Static-noise margin analysis of MOS SRAM cells, IEEE J. Solid State Circuits 22 (5) (1987) 748–754, http://dx.doi.org/10.1109/JSSC.1987.1052809.

4

BTI Impact on the Memory Write-Path

The content of this chapter incorporates the following research article:

1. ***I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, BTI analysis of SRAM write driver, 2015 10th International Design Test Symposium (IDT), pp. 100–105, Dec. 2015, Dead Sea, Jordan.*

2015 10th International Design & Test Symposium (IDT)

BTI Analysis of SRAM Write Driver

Innocent Agbo Mottaqiallah Taouil Said Hamdioui
Delft University of Technology
Faculty of Electrical Engineering, Mathematics and CS
Mekelweg 4, 2628 CD Delft, The Netherlands
{I.O.Agbo, M.Taouil, S.Hamdioui}@tudelft.nl

Pieter Weckx[1,2] Stefan Cosemans[1] Francky Catthoor[1,2]
[1]imec vzw., Kapeldreef 75, B-3001, Leuven, Belgium
[2]Katholieke Universiteit Leuven, ESAT, Belgium
{Pieter.Weckx, stefan.cosemans, Francky.catthoor}@imec.be

***Abstract*—Bias Temperature Instability (BTI) has become a major reliability challenge for nano-scaled devices. This paper presents BTI analysis for the SRAM write driver. Its evaluation metric, the write delay (WD), is analyzed for various supply voltages and temperatures for three technology nodes, i.e., 45nm, 32nm, and 22nm. The results show that as technology scales down, BTI impact on write delay (i.e., its average and $+/-3\sigma$ variations) increases; the 22nm design can degrade up to $1.9\times$ more than the 45nm design at nominal operation conditions. In addition, the result shows that an increment in supply voltage (i.e., from $-10\%V_{dd}$ to $+10\%V_{dd}$) increases the relative write delay during the operational lifetime. Furthermore, the results show that a temperature increment accelerates the BTI induced write delay significantly; while at 298K the degradation is up to 4.7%, it increases to 41.4% at 398K for the 22nm technology node.**

***Index Terms*—BTI, NBTI, PBTI, SRAM write driver**

I. INTRODUCTION

In recent decades, CMOS technology has witnessed relentless downscaling [1]. Forces behind the trend are advancements in the fabrication technology, introduction of novel materials and evolution of architecture designs. However, for circuits based on the current CMOS technology, reliability failures have become a major bottleneck [1]–[3]. Bias Temperature Instability (BTI) (i.e., Negative BTI in PMOS transistors and Positive BTI in NMOS transistors) is a reliability failure mechanism which affects the strength of MOS transistors by increasing their threshold voltages and reducing their drain current (I_d) over the operational lifetime [4]–[6].

Static Random-Access Memories (SRAM) occupy a large part of semiconductor systems and play a major role in the silicon area, performance, and critical robustness [7]. An SRAM system consists of cells array, and its peripherals circuits such as column and row address decoders, control circuits, sense amplifiers and write drivers. Many previous work focused on the BTI SRAM cell array [8] and only a few on the SRAM peripheral circuitry. For example, Binjie *et al.* [9] investigated the NBTI impact on the degradation of Static Noise Margin (SNM) and Write Noise Margin (WNM) for the 6T SRAM cell. Kumar *et al.* [10] analyzed the impact of NBTI on the read stability and SNM of SRAM cells. Bansal *et al.* [11] analyzed the stability of an SRAM cell for worst-case conditions and in the presence of NBTI and PBTI, both individually and combined. Weckx *et al.* [12] investigated a novel statistical based approach for workload dependent circuit lifetime projections which focused on the degradation of SNM metric for 6T SRAM cell. On the other hand, few authors have focused on the reliability analysis of the periheral circuits. Khan *et al.* [13] investigated the combined impact of partial opens and BTI in SRAM address decoder. Menchaca *et al.* [14] analyzed the BTI impact on different sense amplifier designs implemented in 32nm technology node by using failure probability (i.e., flipping a wrong value) as a reliability metric. Agbo *et al.* [15] investigated the BTI impact on SRAM drain-input latch type sense amplifier. In [16] the same authors investigated the integral impact of BTI and voltage temperature variations on SRAM Sense Amplifier. However, to our best knowledge, no previous work focused on the BTI impact on the write driver.

This paper focuses on the write driver design as it has the vital role to ensure correct data is written to the cell [17]. Although, it is obvious that BTI increases delay and reduces memory reliability, the actual BTI impact on the write driver still needs to be characterized. In this regard, the main contributions of the paper are:

- Investigation of BTI impact on the write driver's mean write delay and its distribution.
- Analysis of BTI impact on the write driver at nominal supply voltage for various technology nodes i.e., 45nm, 32nm, and 22nm, respectively; both nominal and high temperature (i.e., 398K) are considered.
- Investigation of BTI impact for varying supply voltages for the write driver design.
- Exploring BTI impact for varying temperatures for the write driver's write delay.

The results depict that with the downscaling of CMOS technology, BTI impact on the write delay causes a significant increment. Moreover, a temperature increase may lead to alarming write delay increment; the impact must be properly investigated to ensure robust and reliable write drivers.

The rest of the paper is organized as follows. Section II introduces the BTI model and memory write driver. Section III provides our analysis and simulation framework, and presents the performed experiments. Section IV analyzes the result for different technology nodes and various supply voltages and temperatures. Finally, Section V concludes the paper

II. BACKGROUND

This section explains first the BTI mechanism. Afterwards, it explains the functional model of an SRAM system. Finally, it presents the behavior of the precharge circuit, SRAM 6T

978-1-4673-9994-4/15/$31.00 ©2015 IEEE

2015 10th International Design & Test Symposium (IDT)

cell, and the analyzed write driver; the precharge and 6T cell are part of the write operation.

A. Bias Temperature Instability

The Bias Temperature Instability (BTI) mechanism takes place inside the MOS transistors and causes a threshold voltage shift that impacts the delay negatively; its mechanism is described below.

BTI Mechanism

BTI increases the absolute V_{th} value in MOS transistors. For the PMOS, the negative V_{th} decreases while for the NMOS the V_{th} increases. The increment in V_{th} in a PMOS transistor that occurs under *negative* gate stress is referred to as NBTI, and the decrement that occurs in an NMOS transistor under *positive* gate stress is known as PBTI. For both MOS transistors, there are two BTI phases, i.e., the stress phase and the relaxation phase.

Recently, exhaustive efforts have been put to understand NBTI [4], [5], [13], [18]. Kaczer *et al.* in [18] have analyzed NBTI using an atomistic model. Alam *et al.* [4] have modeled NBTI and presented the overall dynamics of NBTI as a reaction diffusion process. The model is usable at a higher level such as circuit level. In Kukner *et al.* in [17], both the atomistic and RD models have been compared. In this work, we use the atomistic model as its BTI predictions are more accurate than these from the RD model [17]. We will briefly describe the atomistic model next.

Atomistic Model

Kaczer et al. proposed the atomistic model in [19]. It is based on the capture and emission of single traps during stress and relaxation phases of NBTI/PBTI respectively. The threshold voltage shift of the device ΔV_{th} is the accumulated results of all the capture and emission of carriers in gate oxide defect traps. The probabilities of the defect occupancy in case of capture P_C and emission P_E are defined by

$$P_C(t_{STRESS}) = \frac{\tau_e}{\tau_c+\tau_e}\left\{1-exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{STRESS}\right]\right\} \quad (1)$$

$$P_E(t_{RELAX}) = \frac{\tau_c}{\tau_c+\tau_e}\left\{1-exp\left[-(\frac{1}{\tau_e}+\frac{1}{\tau_c})t_{RELAX}\right]\right\} \quad (2)$$

where τ_c and τ_e are the mean capture and emission time constants, and t_{STRESS} and t_{RELAX} are the stress and relaxation periods, respectively. Furthermore, BTI induced V_{th} is an integral function of Capture/Emission Time CET map [20], workloads, duty factor and transistor dimensions, which gives the mean number of available traps in each device.

B. Memory model

A memory system comprises memory cell array, row and column address decoders, read/write circuitry, input/output data registers and control logic as depicted in functional model of SRAM system in Fig. 1. [21]. The main focus of this paper is the SRAM write driver; in addition, the precharge and SRAM circuit will be also explained as they effect the write operation.

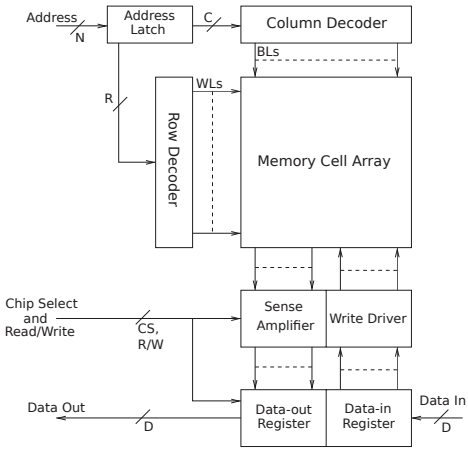


Fig. 1. Functional model of SRAM system

Figure 2 shows the simulation model which is divided into three parts (i.e., precharge circuitry, 6T cell, and the write driver). These parts will be explained below. To be able to read correct data (i.e., either a one or zero) from a memory cell, first valid data must be written into the cell. The circuit that is responsible for correct writing of data into the memory cell is called the write driver. A possible implementation of the write driver is shown in the bottom of Fig. 2. The write operation starts with the precharge circuit that precharges both bit-lines (i.e., *Bl* and *BLBar*). Subsequently, the write driver forces one of the bit-lines to zero, while maintaining the precharged value on the other bit-line. After sometime, when the word line of the cell is activated, the cell's internal state takes over the written values on BL and BLBar.

Precharge circuit

The precharge consists of three PMOS transistors connected to the bit lines (i.e., *BL* and *BLBar*). The differential bit lines are connected to the cell and these bit lines are precharged to high state in typical high-performance memory.

SRAM 6T cell

The SRAM 6T cell consists of six transistors (i.e., two pull-up (PMOS), two pull-down (NMOS), and two pass (NMOS) transistors) which is the basic building block of any memory cell. An SRAM cell has a bistable circuit characteristics which implies that it can be driven into one of the two states and still retains its state as the power goes on.

Write driver

The write driver [21] consists of two pass transistors (i.e., NMOS) and two inverter circuitry. The pass transistors are the transfer device of an SRAM cell, which effectively drives a logic zero but not a logic one. The write operation is realized by writing a zero into either BL or $BLBar$. Thereafter, the

2015 10th International Design & Test Symposium (IDT)

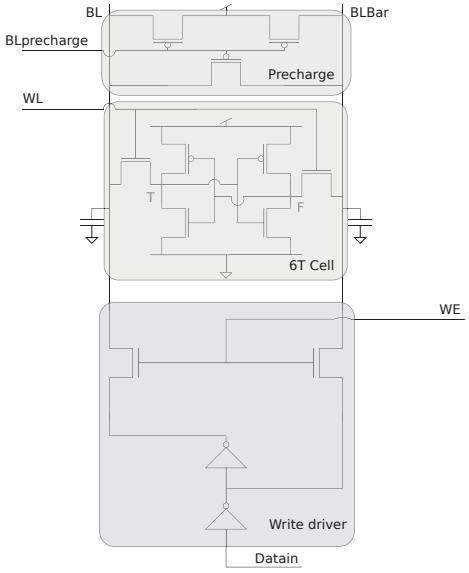


Fig. 2. Simulation setup.

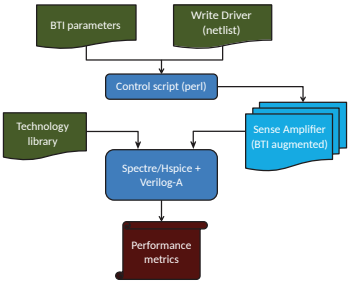


Fig. 3. Analysis framework.

cross-coupled inverters of the cell change the logic zero to a logic one at the other storage node. Then, when the write enable (i.e., WE) signal is activated, the data is transfered to the BLs. Since, forcing a strong zero into one of the BLs is the main target, the PMOS and NMOS devices may be equally sized.

III. Analysis Framework

In this section, the analysis framework and the conducted experiments are presented.

A. Framework Flow

Figure 3 depicts our generic framework to evaluate the BTI impact on the considered memory write driver circuitry. Next, its inputs, processing and output blocks are described.

Input: The general input blocks of the framework are the technology library, memory write driver design, and BTI input parameters. They are explained as follows.

- Technology library: In this work we use three technology nodes; they are 45nm, 32nm, and 22nm PTM library [22]. Note that in general any library card can be used.
- WD design: Generally, all write driver design can be used. In this paper we focus only on the typical write driver (WD) shown in fig. 2. and is described by a SPICE netlist.
- BTI parameters: The BTI induced degradation depends strongly on the stress time duration. The stress time defines how long the workload sequence is being applied. The workload sequence is assumed to be replicated until the age time is reached. The workload assumed in this experiment is based on a 50% duty factor for each transistor of the WD.
- TV: This block specifies the temperature and voltage; each is explained next.

A. Temperature

Temperature variation is an important factor in BTI induced degradation of SRAM write drivers write delay. For instance, MOS transistors threshold voltage reduces as temperature increases. In this paper, we restrict ourselves to temperatures Nominal T_1 = 298K, T_2 = 348K, and T_3 = 398K.

B. Voltage

Supply voltage variations can impact the write delay significantly as it impacts the operational speed. In addition, variation in supply voltage also affects the oxide field (capacitance) inside the transistors and subsequently the BTI impact. In this paper, we restrict ourselves to three supply voltages: V_1 = -10%V_{dd}, V_2 = Nominal V_{dd}, and V_3 = = +10%V_{dd}. Note that the nominal V_{dd} for 45nm is 1.0V while 0.9V for 32nm and 0.8V for 22nm technology.

Processing:

Based on the transistor dimensions and the other specified inputs, the Control script (perl) generates several instances of the BTI augmented SRAM write driver. Every generated instance has a distinct number of traps (with their unique timing constants) in each transistor, and are incorporated in a Verilog-A module of the WD netlist. The module responds to the every individual trap, and alters the transistors concerned parameters such as Vth. After inserting BTI in every transistor of the WD design, a Monte Carlo (MC) is performed at different time steps (100 runs at each time step) where circuit simulator (HSPICE/Spectre) is used to measure the BTI impact. In this paper, our analysis focus on both the mean write delay (denoted by diamond marker in the figure) and its distribution (i.e., +/-3σ denoted by the edges of the vertical lines in the figure).

Output: Finally, statistical post-analysis of the results are performed for varying supply voltages and temperatures in MATLAB environment. The raw outputs are measured directly

2015 10th International Design & Test Symposium (IDT)

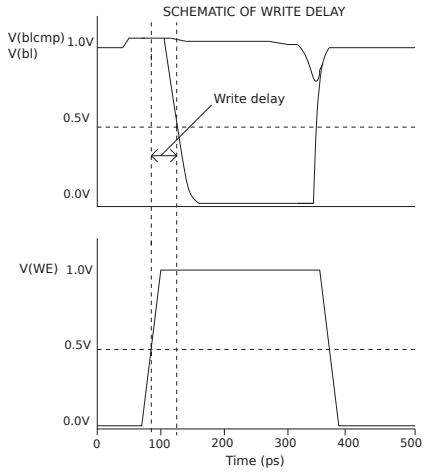


Fig. 4. Metric diagram of Write delay.

from HSPICE/Spectre and measure the write delay. The write delay metric is determined when the trigger signal (i.e., write enable input signal) reaches 50% of the supply voltage and the target (i.e., either bl or $blcmp$ falling output signal) reaches 50% of the supply voltage while the initial conditions of the cell flips to the correct state of the Bit lines. The difference between the target and the trigger results in write delay as shown in Fig. 4. Furthermore, the impact of the BTI degradation is measured relatively to the case where no BTI is present.

B. Experiments Performed

In this paper, four sets of experiments are performed. These experiments are described below:

1. BTI Impact Experiments: The BTI impact on the write delay is investigated for different lifetime.

2. Technology Dependent Experiments: The BTI impact on the write delay for three technology nodes at nominal voltage (i.e., 22nm, 32nm, and 45nm) is investigated. This is performed at nominal and high temperature (i.e., 398K).

3. Supply voltage Dependent Experiments: The BTI impact on the write delay for various supply voltages (i.e., from -10% to +10% V_{dd}) is explored.

4. Temperature Dependent Experiments: The BTI impact on the write delay for various temperatures (i.e., 298K, 348K and 398K) is investigated.

IV. EXPERIMENTAL RESULTS

In this section, we present the analysis results of the experiments mentioned in the previous section.

A. BTI Impact Experiments

Figure 5 shows the absolute mean write delay (denoted by diamond markers) and its distribution, i.e., $+/-3\sigma$ (denoted by the edges of the vertical lines) w.r.t. aging up to 3 years degradation (10^8s) for the write driver.

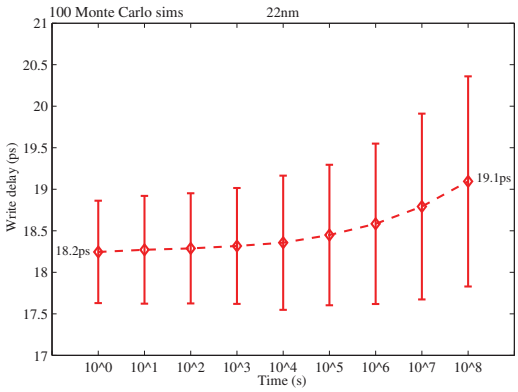


Fig. 5. BTI impact on Write delay for 22nm technology.

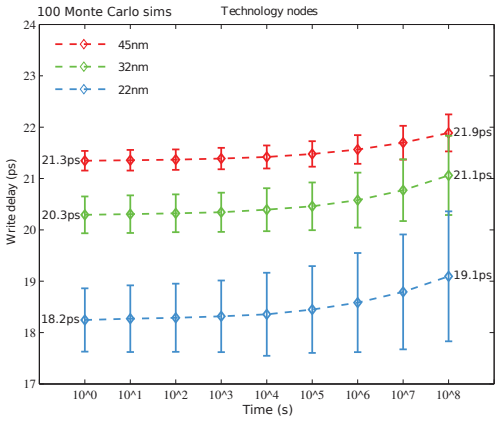


Fig. 6. BTI impact on Write delay for various technology nodes.

The figure shows that the write delay increases over time. For example, the delay increases up to 4.7% in 3 years. The figure also shows that the degradation of the distribution widens over time. For example, the write delay $+3\sigma$ variation increases from 0.6ps to 1.3ps in 3 years.

B. Technology Dependent Experiments

Figure 6 depicts the absolute BTI induced write delay (i.e., its mean and distribution) for the three technology nodes (i.e., 45nm, 32nm, and 22nm) at 298K. The figure shows that all the technology nodes follow the same trends. The smaller the technology node, the higher the relative impact and the wider the $+3\sigma$ variation. For example, after 3 years operational lifetime, the BTI induced relative degradation is 2.82% for 45nm while 3.94% for 32nm node and only 4.95% for 22nm node. In addition, the figure shows that the write delay after 10^8s for 22nm is the fastest when compared to 32nm and

2015 10th International Design & Test Symposium (IDT)

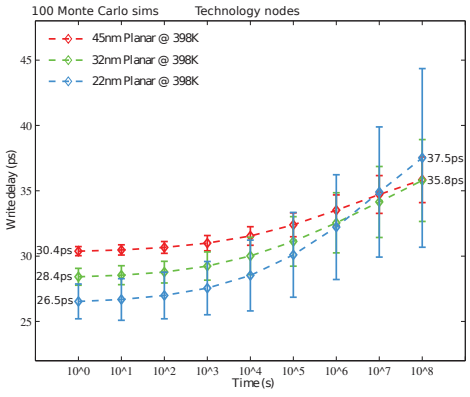


Fig. 7. BTI impact on Write delay for various technology nodes at 398K.

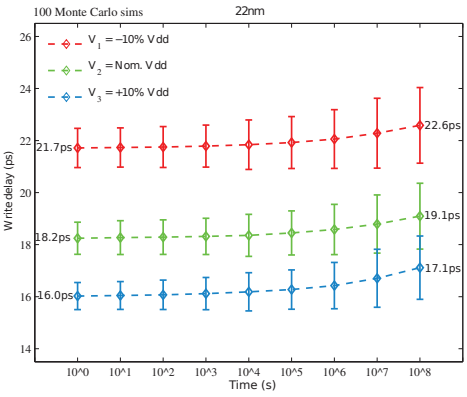


Fig. 8. BTI impact on write delay for varying supply voltages.

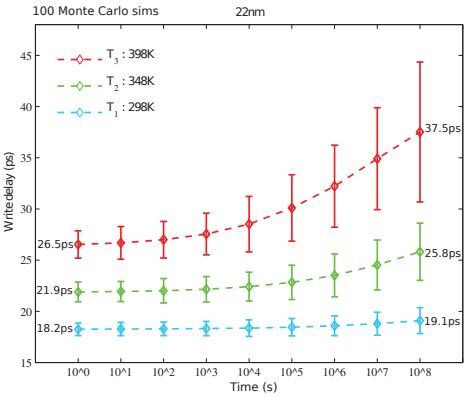


Fig. 9. BTI impact on write delay for varying temperatures.

45nm nodes. For example, after 3 years, the absolute delay degradation for 22nm is up to 1.11× faster than 32nm and only 1.15× faster than 45nm. The figure also shows that the distribution of the degradation is the widest for the lower nodes irrespective of the operational time. For example, after an operation of 10^8s, the BTI induced distribution (i.e., +3σ variation) is 1.3ps for 22nm while 0.8ps and 0.4ps for 32nm and 45nm, respectively. More importantly, the figure shows that after 10^7s the worst case, 22nm drivers (i.e., the slowest ones) become even slower than the worst case drivers of 32nm and 45nm.

Figure 7 shows the BTI induced degradation for different technology nodes at 398K. The figure also shows that the mean BTI induced write delay for 22nm nodes overlaps 32nm and 45nm nodes after an operation of 10^6s to 10^7s; the mean write delay is 5% higher for 22nm at 10^8s over 32nm and 45nm. In addition, the same trend is observed for write delay distribution (i.e., +3σ variation) at 398K. Hence, the degradation distribution is the widest for lower node (i.e., 22nm) at higher temperature.

C. Supply Voltage Dependent Experiments

Figure 8 shows the absolute BTI induced write delay (i.e., average and +/-3σ distribution) for various supply voltages for the write driver design. At 10^0s, when the devices are still fresh, the write driver operates faster at higher V_{dd}. For example, after an operation of 10^0s the absolute write delay is 16.0ps for +10% V_{dd}, and 18.2ps for nominal V_{dd} while 21.7ps for -10% V_{dd}. The figure also shows that the BTI induced write delay increases irrespective of supply voltage considered. For example, after an operation of 10^0s, the degradation is 21.7ps and 22.6ps at 10^8s for -10% V_{dd}, while for 10^0s is 18.2ps and 19.1ps at 10^8s for nominal V_{dd}, and only 16.0ps at 10^0s and 17.1ps at 10^8s for +10% V_{dd}, which shows % increment of 4.0% for -10%V_{dd}, while 4.7% and only 6.8% for nominal V_{dd} and +10%V_{dd}, respectively. It is worth noting that the increase in supply voltage reduces the BTI induced write delay in absolute terms while increases the degradation in relative terms.

The figure also shows that the distribution of the degradation (i.e., +3σ variation) widens as the supply voltage reduces from +10% V_{dd} to -10% V_{dd}. For example, after an operation of 10^8s the +3σ spread equals 1.5ps for +10%V_{dd} and 1.3ps for nominal V_{dd} while 1.2ps for -10%V_{dd}; at +10% V_{dd} this is 1.25× higher than at -10% V_{dd}, while at nominal V_{dd} is 1.08× higher than at -10% V_{dd}.

D. Temperature Dependent Experiments

Figure 9 shows the absolute BTI induced degradation (i.e., average and +/-3σ distribution) for three temperatures (i.e., T_1 = 298K, T_2 = 348K, and T_3 = 398K) for the write driver design. The figure shows that an increase in temperature increases the BTI induced write delay. For example, after an operation of 10^0s the absolute write delay degradation

2015 10th International Design & Test Symposium (IDT)

is 18.2ps for 298K and 21.9ps for 348K while 26.5ps for 398K. Furthermore, after an operation of 10^8s the BTI induced degradation is 19.1ps for 298K and 25.8ps for 348K while 37.5ps for 398K which is about 4.7% for 298K and 17.9% for 348K while 41.4% for 398K.

The figure also shows that the degradation distribution (i.e., +3σ) widens as the temperature increases from 298K to 398K. For example, after an operation of 10^8s the +3σ spread equals 1.2ps for 298K and 2.1ps for 348K, while 3.6ps for 398K; at 398K this is 3$\times$ higher than at 298K and at 348K only 1.75$\times$ higher than at 298K.

E. Discussion

Memory write driver robustness and reliability are very vital for the overall design of memory systems. The presented analysis show that the BTI induced write delay and its distribution of the write driver design is a function of technology, supply voltage, and temperature etc. Evaluating the simulation results with respect to degradation and its variations (i.e., +/-3σ) we conclude the following:

- At lower nodes (i.e., 22nm) the write delay degradation is the worst when compared to other technologies reported in this work; this indicates that BTI may be a serious concern and it may lead to read failures at a lower technology nodes. Moreover, the degradation variation at both corners (i.e., $+/-3\sigma$) of the Planar node is the widest for the 22nm irrespective of the operational lifetime considered, when compared with higher technology nodes. This variation is very crucial and require urgent mitigation technique to address it.
- The degradation increases with decrease in power supply in absolute terms irrespective of the operational lifetime and the distribution follows the same path as the degradation; this implies that at higher power supply, the degradation reduces in absolute terms. However, the mean degradation follows an opposite path in relative terms while the distribution remains the same.
- The degradation and its distribution are worst at a higher temperature (i.e., 398K); this is applied to both absolute and relative terms. The degradation is up to 37.5ps while its distribution is up to 6.8ps at 10^8s operational time.
- The degradation analysis of the memory sub-circuits (i.e., cell, sense amplifier (SA), write driver, etc.) show that SA degrades the most i.e., up to 2.72$\times$ more than the write driver while 16.2$\times$ more than the cell using various delay metrics i.e., sensing delay for SA, and swing delay for the cell [23]; while write delay for the write driver presented in this work.
- The analysis of BTI on write driver design necessitate understanding and quantifying the aging rate for different parts of memory systems for optimal and reliable SRAM systems.

V. CONCLUSION

This paper investigates the impact of Bias Temperature Instability (BTI) for memory write driver for different technologies while considering various supply voltages and temperatures. First, BTI impact increases the write delay and its distributions as the technology nodes scales down causing the memory write driver to be less reliable and robust. Second, increase in supply voltage per technology node compensate the write delay and its variation in absolute terms causing the write driver to be more robust and reliable while this is not the case when viewed from the relative terms point of view. Third, increase in temperature increases the degradation and widens its distribution. These results are validated with HSPICE/Spectre.

REFERENCES

[1] ITRS, "International Technology Roadmap for Semiconductor 2004" "www.itrs.net/common/2004 update/2004update.htm.".

[2] S. Borkar, et al "Micro architecture and Design Challenges for Giga scale Integration", *Pro. of Intl. Sympos. Micro architecture*, 2004.

[3] S. Hamdioui et al, "Reliability Challenges of Real-Time Systems in Forthcoming Technology Nodes", *DATE*, 2013.

[4] M. A. Alam and S. Mahapatra, "A Comprehensive Model of PMOS NBTI Degradation", *Microelectronics Reliability* , Vol:45, 2005.

[5] S. Zafar, Y.H. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, M. Chudzik, "A comparative study of NBTI and PBTI in SiO2/HfO2 stacks with FUSI, TiN gates", *Pro. of VLSI Technology symp.*, 2006.

[6] N. Kizmuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, T. Horiuchi, "The Impact of BTI for Direct Tunneling Ultra Thin Gate Oxide of MOSFET Scaling", *VLSI Technology, Digest of Technical Papers.*, pp: 73-74, 1999.

[7] P. Pouyan et al, "Process Variability-Aware Proactive Reconfiguration Technique for Mitigating Aging effects in Nano Scale SRAM lifetime", *IEEE 30th VLSI Test Symposium.*, 2012.

[8] P. Weckx et al, "Implications of BTI-Induced Time-Dependent Statistics on Yield Estimation of Digital Circuits", *in Electron Devices, IEEE Transactions on , vol.61, no.3, pp.666-673.*, Mar. 2014.

[9] B. Cheng, A. R. Brown, "Impact of NBTI/PBTI on SRAM Stability Degradation", *IEEE ELECTRON DEVICES LETTERS*, 2011.

[10] S. Kumar, C.H. Kim, S. Sapatnekar, "Impact of NBTI on SRAM Read Stability and Design for Reliability", *Pro. of ISQED*, pp: 212-128, 2006.

[11] A. Bansal et al, "Impact of NBTI and PBTI on SRAM static/dynamic noise margins and cell failure probability", *J. Microelectronics Reliability*, 2009.

[12] P. Weckx, et al., "Defect-based Methodology for Workload-dependent Circuit Lifetime Projections-Application to SRAM", *IRPS*, April 2013.

[13] S. Khan, M. Taouil, S. Hamdioui, H. Kukner, P. Raghavan, F. Catthoor, "Impact of Partial Resistive Defects and Bias Temperature Defects and Bias Temperature Instability on SRAM Decoder Reliability", *Pro. of 7th IEEE International Design and Test Symposium*, 2012.

[14] R. Menchaca, H. Mahmoodi, "Impact of Transistor Aging Effects on Sense Amplifier Reliability in Nano-Scale CMOS", *13th Int'l Sym. on Quality Electronic Design*, 2012.

[15] I. Agbo et al, "BTI Impact on SRAM Sense Amplifier", *8th IDT*, 16-18 Dec. 2013.

[16] I. Agbo et al, "Integral Impact of BTI and Voltage Temperature Variation on SRAM Sense Amplifier", *Pro. of 33rd IEEE VTS*, 27-29 Apr. 2015.

[17] H. Kukner et al, "Comparison of Reaction-Diffusion and Atomistic Trap-based Models for Logic Gates", *IEEE TDMR*, 2013.

[18] B. Kackzar, et al., "Disorder-Controlled-Kinetics Model NBTI and its Experimental Verification", *Proc. of Intl. Physics Reliability Symp.(IPRS)*, pp: 381-387, 2005.

[19] B. Kaczer, et al., "Atomistic Approach to Variability of Bias Temperature Instability in Circuit Simulation", *Proc. of IRPS*, April, 2011.

[20] S. Khan, et al, Bias Temperature Instability Analysis of FinFET based SRAM cells, *DATE*, 2014.

[21] H. Kukner, "Generic and Orthogonal March Element based Memory BIST Engine", *MSc Thesis TU Delft, The Netherlands*, 2011.

[22] Predictive Technology Model "http://ptm.asu.edu/",

[23] I. Agbo et al, "Comparative BTI Impact for SRAM Cell and Sense Amplifier Designs", *MEDIAN FINALE workshop*, 10 - 11 Nov. 2015.

5

CONCLUSION

5.1 SUMMARY
5.2 FUTURE RESEARCH DIRECTIONS

This chapter summarizes the comprehensive achievements of this dissertation and underlines some future research directions. Section 5.1 presents a summary of the main conclusions presented in this dissertation. Section 5.2 provides the future research directions.

5.1. SUMMARY

Chapter 1, "Introduction", briefly motivates why the study of memory reliability is crucial in nano scale era. It describes transistor scaling; and the way it affects the variability and reliability of transistors. In addition, it briefly classifies various sources of variability such as time-zero and time-dependent variations. It shows that there are various types of time-dependent variability and the most important one is Bias temperature instability (BTI). BTI presents few opportunities such as reduction in static power consumption while it has a lot of drawbacks.

Chapter 2, "Analysis of aging impact on various memory sense amplifiers", discusses the impact of BTI on memory sense amplifier (SA) which covers two well known aging models and several SA designs. First, it presents a comparative study of the impact of reaction diffusion (RD) and Atomistic trap-based BTI models on the memory sense amplifier. It provides the background for the memory sense amplifier, the BTI models, and the analysis framework. Furthermore, it presents the temporal impact of BTI on sensing delay for the two models on SA design for various stress periods, while taking into account various workloads (applications) and supply voltages.

Second, it discusses the impact of BTI on drain-input latch type SA. It presents the background on SA, BTI model, and the analysis framework which covers the sensing delay and sensing voltage metrics used in this work. It also presents the impact of BTI on sensing delay and sensing voltage of the memory sense amplifier while taking into account various technology nodes such as 90, 65, and 45 nm and various supply voltages. The chapter shows the both metrics (i.e., sensing delay and sensing voltage metrics) leading to clear tradeoff between power and robustness.

Third, it presents the characterization of the integral impact of BTI while considering process, voltage, and temperature (PVT) variations on the memory sense amplifier for various technology nodes such as 45, 32, 22, and 16 nm. It also provides background of the standard latch type SA and the RD model. Furthermore, it gives the analysis framework, the workloads, and the performance metric. Moreover, it presents the impact of BTI as well as impact of BTI combined with PVT on sensing delay of the memory sense amplifier, while taking into account various workloads, technology nodes, supply voltages, and temperatures.

Fourth, it proposes an accurate methodology to investigate the impact of both time-zero and time-dependent fluctuations on the offset voltage specification of a memory SA design for 45 nm, and provides thorough flow of the proposed technique for the offset voltage specification of the SA. In addition, it covers the sensitivity analysis of the offset voltage specification for all the process corners; it presents the impact of process variation on the offset voltage. In addition, it presents the impact of combined effect of process and BTI variation on the offset voltage specification, and presents the failure rate analysis for the nominal process corner. Most of the analysis done in this chapter takes the different supply voltages, temperatures, and workloads into consideration. The pro-

posed technique in this chapter provides designers a favorable way of determining the offset voltage specification, and helps them in making adequate design choices dependent on the applications and environmental conditions.

The chapter also discusses a comparative study of the impact of BTI on three memory SAs for 45nm technology node. It presents the framework analysis flow and the SA offset specification technique, and covers the impact of BTI on sensing delay of three SA designs for various supply voltages and temperatures. On top of that, it explores the impact of technology scaling on sensing delay degradation of a high performance SA. The results indicate that designing for reliability is not only strongly workload dependent, but also technology node dependent.

Chapter 3, "Investigation of Read path aging", focuses on the analysis of the impact of BTI on the memory read path. It presents the metrics used for the analysis such as bitline swing and sense amplifier sensing delay for four different combinations of the workloads.

The chapter also proposes an adequate technique to quantify and mitigate the aging on the read path of a high performance memory design for various supply voltages, temperatures, workloads, and technology nodes. On top of that, it presents the impact of BTI on dynamic energy for different device strengths. Finally, it analyzes the impact of BTI on cell stability of HSNM, RSNM, and WTP while considering both nominal cell size and halve cell size.

Chapter 4, "Investigation of Write path aging", focuses on the analysis of the BTI on the memory write driver. Memory write driver is very important to ensure that correct data is written into the cell array. Therefore, understanding, characterization, and analysis of the actual impact of BTI is required. Our analysis takes into account various technologies, supply voltages, and temperatures and it was validated with HSPICE/Spectre simulations.

5.2. FUTURE RESEARCH DIRECTIONS

Several research directions are proposed to further extend some aspects of topics addressed in this dissertation. The research directions are categorized into two main subsections, 5.2.1 and 5.2.2 which present memory aging modeling and mitigation schemes, respectively. They are explained next.

5.2.1. MEMORY AGING MODELING

The scaling of CMOS technology has given rise to reliability issues. But the most researched reliability issue is the Bias Temperature Instability that impacts both logic and memory circuits. However, there are other aging failure mechanisms such as Hot Carrier Injection (HCI), Time Dependent Dielectric Breakdown (TDDB), Electro-Migration (EM) and so on. These failure mechanisms also impact the reliability of devices, inter-

connects, and memory circuits. In addition, the impact of such mechanisms on different memory technologies and designs are not covered in this work. Some future research topics on memory reliability modeling are given next.

- **Consider the impact of other failure mechanisms:** The impact of Bias Temperature Instability on embedded memories has been simulated and analyzed for various experiments while other failure mechanisms are yet to be explored. The simulation and analysis of the impact of HCI, TDDB, and EM failure mechanisms, irrespective of the circuitry can be investigated. Moreover, performing the aging analysis using these models will reveal the extent of their impact to any specific circuit while considering an appropriate mitigation scheme to compensate for these impacts.

- **Impact on other technology nodes and designs:** The degradation of IC may not be only CMOS technology scaling dependent; it may also depend on other process technologies, and designs.

  - Most prior work has analyzed the impact of aging for sub-100 nm technology nodes up to 22 nm while other newer technology nodes such as sub-14 nm and 7 nm technology nodes are yet to be explored. As a result, it is crucial to investigate the aging of these lower technology nodes so as to provide data to the memory system designers.
  - Prior work mostly has focused on Planar CMOS; limited work exists on FinFET technology nodes and other process technologies such as Silicon-on-Insulator (SOI), Fully depleted Silicon-on-Insulator (FDSOI) and so on; these need more attention especially for smaller nodes. On top of that, below 14 nm, the device stacks and cell as well as periphery circuits are further changing due to the added boosters. That will require also major research effort in the future. Obviously, the dependency on supply voltages fluctuations, temperatures, and workloads have to be investigated. This is important for the memory designer to make vital decisions with respect to other process technology nodes.
  - Exploring the aging of new and/or emerging new memory architecture/design is of equal importance as well. The aging analysis of this dissertation was mainly on the memory systems with a particular focus on 6T cell architecture. The analysis of aging on the new memory design with different cell structures such as 4T, 8T, and so on, while taking into account circuit level structure (e.g., scalability and size), parametric, and data dependency (e.g., application and PVT), need still to be fully explored.

- **Analysis of the whole memory system:** The memory system is made up of the cell array, sense amplifier, write drivers, row and column decoders, control timing circuit, and so on. Therefore, it is crucial to understand the interaction between these components and how they contribute to the overall memory degradation. E.g., that an individual memory component is the most sensitive to aging may not necessarily mean that that component is the major or / most contributor to the overall

memory aging. Hence, the need of the analysis of the impact of different reliability failure mechanisms on the entire memory systems and how the individual memory components interact with each other while taking into account various supply voltages, temperatures and workloads is of great importance for optimal designs.

5.2.2. MITIGATION SCHEMES

The purpose of reliability modeling is to better understand the degradation in order to come up with appropriate mitigation schemes that could extend the lifetime of the memory system and/or reduce the in-field failure rate. There are two major types of mitigation schemes and these includes functional and parametric mitigation schemes. Moreover, functional mitigation technique makes use of adding hardware modules either with same functionality or different functionality while performing parallel execution or switching to spare module when the parallel module fails, (for example, adaptive resource management) while parametric mitigation technique makes use of either static or dynamic adjustment of supply voltage, temperature, frequency margins and etc., so as to compensate the degradation due to Aging. In summary, there are three types of mitigation schemes that could be explored.

- **Static:** The static mitigation approaches compensate for aging using a fixed scheme. For example, device resizing, fixed supply voltage and frequency margins, and balancing paths based on predicted degradation due to aging are commonly used schemes. However, these schemes typically come with an additional cost both in terms of area and performance due to over-design. The results in this dissertation show that the degradation strongly depends on the workload, temperature and applied voltage. Therefore, the question is whether a static scheme is the right approach to take in addressing the degradation of memory systems. One possible approach is to make a trade-off between the different techniques and select different mitigation techniques for the different parts of the memory. Nevertheless, the effectiveness of the static schemes remains limited as they require an accurate aging precision.

- **Dynamic adaptation:** Mitigation techniques based on dynamic adaptation compensate the degradation while taking the severeness of aging into account at run-time. Such schemes consist typically of a monitoring circuit that measures the degradation and an actuator that adjust key system parameters such as operating voltage and frequency. Examples are dynamic voltage and frequency scaling, adaptive body biasing, bit-flipping techniques compensate for aging at run-time. As the degradation strongly depends on the supply voltage, temperature and workload, dynamic adaptation schemes might be the right step to take in addressing the reliability issues faced by the memory system. This requires however an acceptable overhead.

- **Adaptive resource management:** The adaptive resource management mitigation schemes focus on balancing aging based on available resources. These schemes can also be applied to the memory system. Typically, a memory system consists

of multiple smaller arrays. One or multiple of these smaller arrays could be used as a redundant element. By powering down different sub-arrays over time, the overall lifetime of the memory can be extended. Key questions to investigate are the frequency of activating and disabling sub-arrays, how long the recovery should take, and how to maintain data integrity. On top of that, the benefits and cost overhead of such schemes need also to be investigated.

REFERENCES

[1] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, Jan 1998.

[2] S. Hamdioui, D. Gizopoulos, G. Guido *et al.*, "Reliability challenges of real-time systems in forthcoming technology nodes," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2013, pp. 129–134.

[3] G. Groeseneken, R. Degraeve, B. Kaczer *et al.*, "Trends and perspectives for electrical characterization and reliability assessment in advanced cmos technologies," in *2010 Proceedings of the European Solid State Device Research Conference*, Sep 2010, pp. 64–72.

[4] R. H. Dennard, F. H. Gaensslen, V. L. Rideout *et al.*, "Design of ion-implanted mosfet's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct 1974.

[5] R. H. Dennard, F. H. Gaensslen, V. L. Rideout *et al.*, "Design of ion-implanted mosfet's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct 1974.

[6] S. H. Kukner, *Bias Temperature Instability in CMOS Digital Circuits from Planar to FinFET Nodes*. Ph.D Thesis Katholieke Universiteit Leuven, 2015.

[7] J. Franco, B. Kaczer, and G. Groeseneken, *Reliability of High Mobility SiGe Channel MOSFETs for Future CMOS Applications*, 1st ed. Springer Science+Business Media Dordrecht: Springer Netherlands, 2014.

[8] E. Ungersboeck, V. Sverdlov, H. Kosina *et al.*, "Strain engineering for cmos devices," in *2006 8th International Conference on Solid-State and Integrated Circuit Technology Proceedings*, Oct 2006, pp. 124–127.

[9] K. Mistry, C. Allen, C. Auth *et al.*, "A 45nm logic technology with high-k+metal gate transistors, strained silicon, 9 cu interconnect layers, 193 nm dry patterning, and 100% pb-free packaging," in *2007, IEEE International Electron Devices Meeting*, Dec 2007, pp. 247–250.

[10] S. Natarajan, M. Armstrong, M. Bost *et al.*, "A 32nm logic technology featuring 2nd-generation high-k + metal-gate transistors, enhanced channel strain and $0.171\mu m^2$ sram cell size in a 291mb array," in *2008 IEEE International Electron Devices Meeting*, Dec 2008, pp. 1–3.

[11] S. Natarajan, M. Agostinelli, S. Akbar *et al.*, "A 14nm logic technology featuring 2nd-generation finfet, air-gapped interconnects, self-aligned double patterning and a 0.0588μm^2 sram cell size," in *2014 IEEE International Electron Devices Meeting*, Dec 2014, pp. 3.7.1–3.7.3.

[12] C. Auth, C. Allen, A. Blattner *et al.*, "A 22nm high performance and low-power cmos technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density mim capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, June 2012, pp. 131–132.

[13] M. Bohr, "A 30 year retrospective on dennard's mosfet scaling paper," *IEEE Solid-State Circuits Society Newsletter*, vol. 12, no. 1, pp. 11–13, Winter 2007.

[14] P. M. Ferreira, H. Cai, and L. Naviner, "Reliability aware ams/rf performance optimization," in *Performance Optimization Techniques in Analog, Mixed-Signal, and Radio-Frequency Circuit Design*, 2014, pp. 1–26.

[15] T. Mak, "Is cmos more reliable with scaling?" in *CRC IEEE BAST workshop*, 2003, pp. 1–17.

[16] T. Grasser, B. Kaczer, W. Goes *et al.*, "Recent advances in understanding the bias temperature instability," in *2010 International Electron Devices Meeting*, Dec 2010, pp. 4.4.1–4.4.4.

[17] T. Grasser, B. Kaczer, W. Goes *et al.*, "The paradigm shift in understanding the bias temperature instability: From reaction-diffusion to switching oxide traps," *IEEE Transactions on Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov 2011.

[18] B. Kaczer, T. Grasser, P. J. Roussel *et al.*, "Origin of nbti variability in deeply scaled pfets," in *2010 IEEE International Reliability Physics Symposium*, May 2010, pp. 26–32.

[19] X. Wang, A. R. Brown, B. Cheng *et al.*, "Statistical variability and reliability in nanoscale finfets," in *2011 International Electron Devices Meeting*, Dec 2011, pp. 5.4.1–5.4.4.

[20] O. S. Unsal, J. W. Tschanz, K. Bowman *et al.*, "Impact of parameter variations on circuits and microarchitecture," *IEEE Micro*, vol. 26, no. 6, pp. 30–39, Nov 2006.

[21] M. Nourani and A. Radhakrishnan, "Testing on-die process variation in nanometer vlsi," *IEEE Design Test of Computers*, vol. 23, no. 6, pp. 438–451, June 2006.

[22] S. Ghosh and K. Roy, "Parameter variation tolerance and error resiliency: New design paradigm for the nanoscale era," *Proceedings of the IEEE*, vol. 98, no. 10, pp. 1718–1751, Oct 2010.

[23] S. K. Saha, "Modeling process variability in scaled cmos technology," *IEEE Design Test of Computers*, vol. 27, no. 2, pp. 8–16, March 2010.

[24] J. Bastos, M. S. J. Steyaert, A. Pergoot *et al.*, "Influence of die attachment on mos transistor matching," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 2, pp. 209–218, May 1997.

[25] Y. Ohnari, A. A. Khan, A. Dutta *et al.*, "Die-to-die and within-die variation extraction for circuit simulation with surface-potential compact model," in *2013 IEEE International Conference on Microelectronic Test Structures (ICMTS)*, March 2013, pp. 146–150.

[26] P. Yang, D. E. Hocevar, P. F. Cox *et al.*, "An integrated and efficient approach for mos vlsi statistical circuit design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 5, no. 1, pp. 5–14, Jan 1986.

[27] N. Herr and J. J. Barnes, "Statistical circuit simulation modeling of cmos vlsi," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 5, no. 1, pp. 15–22, Jan 1986.

[28] S. Ghosh and K. Roy, "Parameter variation tolerance and error resiliency: New design paradigm for the nanoscale era," *Proceedings of the IEEE*, vol. 98, no. 10, pp. 1718–1751, Oct 2010.

[29] A. Asenov, A. R. Brown, J. H. Davies *et al.*, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale mosfets," *IEEE Transactions on Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep 2003.

[30] A. Asenov, "Simulation of statistical variability in nano mosfets," in *2007 IEEE Symposium on VLSI Technology*, June 2007, pp. 86–87.

[31] P. A. Stolk and D. B. M. Klaassen, "The effect of statistical dopant fluctuations on mos device performance," in *International Electron Devices Meeting. Technical Digest*, Dec 1996, pp. 627–630.

[32] D. J. Frank, Y. Taur, M. Ieong *et al.*, "Monte carlo modeling of threshold variation due to dopant fluctuations," in *1999 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.99CH36325)*, June 1999, pp. 169–170.

[33] H. Mahmoodi, S. Mukhopadhyay, and K. Roy, "Estimation of delay variations due to random-dopant fluctuations in nanoscale cmos circuits," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 9, pp. 1787–1796, Sep 2005.

[34] G. Roy, A. R. Brown, F. Adamu-Lema *et al.*, "Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-mosfets," *IEEE Transactions on Electron Devices*, vol. 53, no. 12, pp. 3063–3070, Dec 2006.

[35] Y. Ye, F. Liu, M. Chen *et al.*, "Statistical modeling and simulation of threshold variation under random dopant fluctuations and line-edge roughness," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 6, pp. 987–996, June 2011.

[36] Y. Li, C. H. Hwang, and T. Y. Li, "Random-dopant-induced variability in nano-cmos devices and digital circuits," *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1588–1597, Aug 2009.

[37] A. R. Brown, N. Daval, K. K. Bourdelle *et al.*, "Comparative simulation analysis of process-induced variability in nanoscale soi and bulk trigate finfets," *IEEE Transactions on Electron Devices*, vol. 60, no. 11, pp. 3611–3617, Nov 2013.

[38] K. J. Kuhn, M. D. Giles, D. Becher *et al.*, "Process technology variation," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug 2011.

[39] K. J. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale cmos," in *2007 IEEE International Electron Devices Meeting*, Dec 2007, pp. 471–474.

[40] P. Oldiges, Q. Lin, K. Petrillo *et al.*, "Modeling line edge roughness effects in sub 100 nanometer gate length devices," in *2000 International Conference on Simulation Semiconductor Processes and Devices (Cat. No.00TH8502)*, 2000, pp. 131–134.

[41] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer mosfets introduced by gate line edge roughness," *IEEE Transactions on Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May 2003.

[42] B. Cheng, A. R. Brown, X. Wang *et al.*, "Statistical variability study of a 10nm gate length soi finfet device," in *2012 IEEE Silicon Nanoelectronics Workshop (SNW)*, June 2012, pp. 1–2.

[43] H.-W. Kim, J.-Y. Lee, J. Shin *et al.*, "Experimental investigation of the impact of lwr on sub-100-nm device performance," *IEEE Transactions on Electron Devices*, vol. 51, no. 12, pp. 1984–1988, Dec 2004.

[44] S. Mori, T. Morisawa, N. Matsuzawa *et al.*, "Reduction of line edge roughness in the top surface imaging process," *Journal of Vacuum Science & Technology B*, vol. 16, no. 6, pp. 3739–3743, Sep 1998.

[45] V. R. Manfrinato, L. Zhang, H. D. D. Su *et al.*, "Resolution limits of electron-beam lithography toward the atomic scale," *Nano Letters*, vol. 13, no. 4, pp. 1555–1558, March 2013.

[46] X. Wang, A. R. Brown, N. Idris *et al.*, "Statistical threshold-voltage variability in scaled decananometer bulk hkmg mosfets: A full-scale 3-d simulation scaling study," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2293–2301, Aug 2011.

[47] A. R. Brown, N. M. Idris, J. R. Watling *et al.*, "Impact of metal gate granularity on threshold voltage variability: A full-scale three-dimensional statistical simulation study," *IEEE Electron Device Letters*, vol. 31, no. 11, pp. 1199–1201, Nov 2010.

[48] A. R. Brown, G. Roy, and A. Asenov, "Poly-si-gate-related variability in decananometer mosfets with conventional architecture," *IEEE Transactions on Electron Devices*, vol. 54, no. 11, pp. 3056–3063, Nov 2007.

[49] A. Asenov and S. Saini, "Polysilicon gate enhancement of the random dopant induced threshold voltage fluctuations in sub-100 nm mosfets with ultrathin gate oxide," *IEEE Transactions on Electron Devices*, vol. 47, no. 4, pp. 805–812, April 2000.

[50] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano mosfets due to local oxide thickness variations," *IEEE Transactions on Electron Devices*, vol. 49, no. 1, pp. 112–119, Jan 2002.

[51] P. Andrei and I. Mayergoyz, "Quantum mechanical effects on random oxide thickness and random doping induced fluctuations in ultrasmall semiconductor devices," *Journal of Applied Physics*, vol. 94, no. 11, pp. 7163–7172, Nov 2003.

[52] M. Niwa, T. Kouzaki, K. Okada *et al.*, "Atomic-order planarization of ultrathin s i o 2 / s i (001) interfaces," *Japanese Journal of Applied Physics*, vol. 33, no. 1S, pp. 388–394, Nov 1994. [Online]. Available: http://stacks.iop.org/1347-4065/33/i=1S/a=388

[53] C. Shin, *Variation-Aware Advanced CMOS Devices and SRAM*. Springer Science+Business Media Dordrecht 2016, 2016.

[54] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. USA: Addison-Wesley Publishing Company, 2010.

[55] ITRS, "International technology roadmap for semiconductor 2013, www.itrs.net/common/2013 update/2013 update.htm," *SIA*, 2013.

[56] S. S. Sapatnekar, "Overcoming variations in nanometer-scale technologies," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 1, no. 1, pp. 5–18, March 2011.

[57] R. Wang, J. Dunkley, T. A. DeMassa *et al.*, "Threshold voltage variations with temperature in mos transistors," *IEEE Transactions on Electron Devices*, vol. 18, no. 6, pp. 386–388, Jun 1971.

[58] T. Karnik and P. Hazucha, "Characterization of soft errors caused by single event upsets in cmos processes," *IEEE Transactions on Dependable and Secure Computing*, vol. 1, no. 2, pp. 128–143, April 2004.

[59] R. C. Baumann, "Soft errors in advanced semiconductor devices-part i: the three radiation sources," *IEEE Transactions on Device and Materials Reliability*, vol. 1, no. 1, pp. 17–22, March 2001.

[60] P. Hazucha, C. Svensson, and S. A. Wender, "Cosmic-ray soft error rate characterization of a standard 0.6-/spl mu/m cmos process," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 10, pp. 1422–1429, Oct 2000.

[61] P. Hazucha and C. Svensson, "Impact of cmos technology scaling on the atmospheric neutron soft error rate," *IEEE Transactions on Nuclear Science*, vol. 47, no. 6, pp. 2586–2594, Dec 2000.

[62] S. P. Park, K. Kang, and K. Roy, "Reliability implications of bias-temperature instability in digital ics," *IEEE Design Test of Computers*, vol. 26, no. 6, pp. 8–17, Nov 2009.

[63] K. O. Jeppson and C. M. Svensson, "Negative bias stress of mos devices at high electric fields and degradation of mnos devices," *Journal of Applied Physics*, vol. 48, no. 5, pp. 2004–2014, Aug 1977.

[64] Y. Miura and Y. Matukura, "Investigation of silicon-silicon dioxide interface using mos structure," *Japanese Journal of Applied Physics*, vol. 5, no. 2, p. 180, 1966. [Online]. Available: http://stacks.iop.org/1347-4065/5/i=2/a=180

[65] B. E. Deal, M. Sklar, A. S. Grove *et al.*, "Characteristics of the surfacestate charge (qss) of thermally oxidized silicon," *Journal of The Electrochemical Society*, vol. 114, no. 3, pp. 266–274, 1967.

[66] D. Frohman-Bentchkowsky, "A fully decoded 2048-bit electrically programmable famos read-only memory," *IEEE Journal of Solid-State Circuits*, vol. 6, no. 5, pp. 301–306, Oct 1971.

[67] W. Wang, S. Yang, S. Bhardwaj *et al.*, "The impact of nbti effect on combinational circuit: Modeling, simulation, and analysis," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 2, pp. 173–183, Feb 2010.

[68] N. Kimizuka, T. Yamamoto, T. Mogami *et al.*, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on mosfet scaling," in *1999 Symposium on VLSI Technology. Digest of Technical Papers (IEEE Cat. No.99CH36325)*, June 1999, pp. 73–74.

[69] K. Kang, S. P. Park, K. Roy *et al.*, "Estimation of statistical variation in temporal nbti degradation and its impact on lifetime circuit performance," in *Proceedings of the 2007 IEEE/ACM International Conference on Computer-aided Design*, ser. ICCAD '07, 2007, pp. 730–734. [Online]. Available: http://dl.acm.org/citation.cfm?id=1326073.1326227

[70] S. P. Park, K. Kang, and K. Roy, "Reliability implications of bias-temperature instability in digital ics," *IEEE Design Test of Computers*, vol. 26, no. 6, pp. 8–17, Nov 2009.

[71] B. Kaczer, V. Arkhipov, R. Degraeve *et al.*, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *2005 IEEE International Reliability Physics Symposium, 2005. Proceedings. 43rd Annual.*, April 2005, pp. 381–387.

[72] V. Huard, M. Denais, and C. Parthasarathy, "Nbti degradation: From physical mechanisms to modelling," *Microelectronics Reliability*, vol. 46, no. 1, pp. 1–23, Jan 2006.

[73] M. Cho, J. D. Lee, M. Aoulaiche *et al.*, "Insight into n/pbti mechanisms in sub-1-nm-eot devices," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2042–2048, Aug 2012.

[74] M. A. Alam, "A critical examination of the mechanics of dynamic nbti for pmosfets," in *IEEE International Electron Devices Meeting 2003*, Dec 2003, pp. 14.4.1–14.4.4.

[75] A. T. Krishnan, C. Chancellor, S. Chakravarthi *et al.*, "Material dependence of hydrogen diffusion: implications for nbti degradation," in *IEEE InternationalElectron Devices Meeting, 2005. IEDM Technical Digest.*, Dec 2005, pp. 4 pp.–691.

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

[77] S. Chakravarthi, A. Krishnan, V. Reddy *et al.*, "A comprehensive framework for predictive modeling of negative bias temperature instability," in *2004 IEEE International Reliability Physics Symposium. Proceedings.*, April 2004, pp. 273–282.

[78] S. Mahapatra, M. Alam, P. B. Kumar *et al.*, "Negative bias temperature instability in CMOS devices," *Microelectronic Engineering*, vol. 80, no. 0, pp. 114–121, June 2005.

[79] H. Kukner, P. Weckx, J. Franco *et al.*, "Scaling of bti reliability in presence of time-zero variability," in *2014 IEEE International Reliability Physics Symposium*, June 2014, pp. CA.5.1–CA.5.7.

[80] V. Reddy, J. Carulli, A. Krishnan *et al.*, "Impact of negative bias temperature instability on product parametric drift," in *2004 International Conferce on Test*, Oct 2004, pp. 148–155.

[81] M. Houssa, S. D. Gendt, J. L. Autran *et al.*, "Detrimental impact of hydrogen on negative bias temperature instabilities in hfo2-based pmosfets," in *Digest of Technical Papers. 2004 Symposium on VLSI Technology, 2004.*, June 2004, pp. 212–213.

[82] R. Fernandez, B. Kaczer, A. Nackaerts *et al.*, "Ac nbti studied in the 1 hz – 2 ghz range on dedicated on-chip cmos circuits," in *2006 International Electron Devices Meeting*, Dec 2006, pp. 1–4.

[83] M. B. Ketchen, M. Bhushan, and R. Bolam, "Ring oscillator based test structure for nbti analysis," in *2007 IEEE International Conference on Microelectronic Test Structures*, March 2007, pp. 42–47.

[84] K. Hofmann, H. Reisinger, K. Ermisch *et al.*, "Highly accurate product-level aging monitoring in 40nm cmos," in *2010 Symposium on VLSI Technology*, June 2010, pp. 27–28.

[85] M. Toledano-Luque, B. Kaczer, P. J. Roussel *et al.*, "Response of a single trap to ac negative bias temperature stress," in *2011 International Reliability Physics Symposium*, April 2011, pp. 4A.2.1–4A.2.8.

[86] J. J. Kim, B. P. Linder, R. M. Rao *et al.*, "Reliability monitoring ring oscillator structures for isolated/combined nbti and pbti measurement in high-k metal gate technologies," in *2011 International Reliability Physics Symposium*, April 2011, pp. 2B.4.1–2B.4.4.

[87] S. Khan, I. Agbo, S. Hamdioui *et al.*, "Bias temperature instability analysis of finfet based sram cells," in *Proceedings of the Conference on Design, Automation & Test in Europe*, ser. DATE '14, March 2014, pp. 31:1–31:6.

[88] P. Weckx, B. Kaczer, M. Toledano-Luque *et al.*, "Defect-based methodology for workload-dependent circuit lifetime projections - application to sram," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, April 2013, pp. 3A.4.1–3A.4.7.

[89] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "An analytical model for negative bias temperature instability," in *Proceedings of the 2006 IEEE/ACM International Conference on Computer-aided Design*, ser. ICCAD '06. New York, NY, USA: ACM, 2006, pp. 493–496. [Online]. Available: http://doi.acm.org/10.1145/1233501.1233601

[90] T. Grasser, B. Kaczer, W. Goes *et al.*, "The paradigm shift in understanding the bias temperature instability: From reaction 2013;diffusion to switching oxide traps," *IEEE Transactions on Electron Devices*, vol. 58, no. 11, pp. 3652–3666, Nov 2011.

[91] B. Kaczer, T. Grasser, J. Roussel *et al.*, "Ubiquitous relaxation in bti stressing 2014;new evaluation and insights," in *2008 IEEE International Reliability Physics Symposium*, April 2008, pp. 20–27.

[92] B. Vaidyanathan and A. S. Oates, "Technology scaling effect on the relative impact of nbti and process variation on the reliability of digital circuits," *IEEE Transactions on Device and Materials Reliability*, vol. 12, no. 2, pp. 428–436, June 2012.

[93] S. Zafar, A. Callegari, E. Gusev *et al.*, "Charge trapping in high k gate dielectric stacks," in *Digest. International Electron Devices Meeting,*, Dec 2002, pp. 517–520.

[94] A. Shanware, M. R. Visokay, J. J. Chambers *et al.*, "Characterization and comparison of the charge trapping in hfsion and hfo/sub 2/ gate dielectrics," in *IEEE International Electron Devices Meeting 2003*, Dec 2003, pp. 38.6.1–38.6.4.

[95] S. Zafar, Y. Kim, V. Narayanan *et al.*, "A comparative study of nbti and pbti (charge trapping) in sio2/hfo2 stacks with fusi, tin, re gates," in *2006 Symposium on VLSI Technology, 2006. Digest of Technical Papers.*, 2006, pp. 23–25.

[96] J. Tschanz, K. Bowman, S. Walstra *et al.*, "Tunable replica circuits and adaptive voltage-frequency techniques for dynamic voltage, temperature, and aging variation tolerance," in *2009 Symposium on VLSI Circuits*, June 2009, pp. 112–113.

[97] K. T. Lee, W. Kang, E. A. Chung *et al.*, "Technology scaling on high-k amp; metal-gate finfet bti reliability," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, April 2013, pp. 2D.1.1–2D.1.4.

[98] S. Pae, M. Agostinelli, M. Brazier *et al.*, "Bti reliability of 45 nm high-k + metal-gate process technology," in *2008 IEEE International Reliability Physics Symposium*, April 2008, pp. 352–357.

[99] E. Mintarno, V. Chandra, D. Pietromonaco *et al.*, "Workload dependent nbti and pbti analysis for a sub-45nm commercial microprocessor," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, April 2013, pp. 3A.1.1–3A.1.6.

[100] V. Huard, E. Pion, F. Cacho *et al.*, "A predictive bottom-up hierarchical approach to digital system reliability," in *2012 IEEE International Reliability Physics Symposium (IRPS)*, April 2012, pp. 4B.1.1–4B.1.10.

[101] H. Kükner, M. Khatib, S. Morrison *et al.*, "Degradation analysis of datapath logic subblocks under nbti aging in finfet technology," in *Fifteenth International Symposium on Quality Electronic Design*, March 2014, pp. 473–479.

[102] T. T.-H. Kim and Z. H. Kong, "Impact analysis of nbti/pbti on sram vmin and design techniques for improved sram vmin," *JSTS: Journal of Semiconductor Technology and Science*, vol. 13, no. 2, pp. 87–97, April 2013.

[103] Z. Chen, K. Hess, J. Lee *et al.*, "On the mechanism for interface trap generation in mos transistors due to channel hot carrier stressing," *IEEE Electron Device Letters*, vol. 21, no. 1, pp. 24–26, Jan 2000.

[104] M. White and J. B. Bernstein, "Microelectronics reliability : physics-of-failure based modeling and lifetime evaluation," *tech. rep., Jet Propulsion Laboratory NASA*, Feb 2008.

[105] S. Mahapatra, D. Saha, D. Varghese *et al.*, "On the generation and recovery of interface traps in mosfets subjected to nbti, fn, and hci stress," *IEEE Transactions on Electron Devices*, vol. 53, no. 7, pp. 1583–1592, July 2006.

[106] R. E. Corporation, "Semiconductor reliability handbook," *rev.0.5 ed.*, Sep 2010.

[107] J. B. Bernstein, M. Gurfinkel, X. Li *et al.*, "Electronic circuit reliability modeling," *Microelectronics Reliability*, vol. 46, no. 7, pp. 1957–1979, July 2006.

[108] X. Li, J. Qin, and J. B. Bernstein, "Compact modeling of mosfet wearout mechanisms for circuit-reliability simulation," *IEEE Transactions on Device and Materials Reliability*, vol. 8, no. 1, pp. 98–121, March 2008.

[109] R. Degraeve, G. Groeseneken, R. Bellens *et al.*, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 904–911, April 1998.

[110] A. T. Krishnan and P. E. Nicollian, "Analytic extension of the cell-based oxide breakdown model to full percolation and its implications," in *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, April 2007, pp. 232–239.

[111] D. Bergstrom, M. Hattendorf, J. Hicks *et al.*, "45nm transistor reliability," *Intel Technology Journal*, vol. 12, no. 2, pp. 131–144, June 2008.

[112] J. H. Stathis, "Physical and predictive models of ultrathin oxide reliability in cmos devices and circuits," *IEEE Transactions on Device and Materials Reliability*, vol. 1, no. 1, pp. 43–59, March 2001.

[113] L. Zhao, Z. Tőkei, K. Croes *et al.*, "Direct observation of the 1/e dependence of time dependent dielectric breakdown in the presence of copper," *Applied Physics Letters*, vol. 98, no. 3, Jan 2011.

[114] T. K. Wong, "Time dependent dielectric breakdown in copper low-k interconnects: Mechanisms and reliability models," *Materials*, vol. 5, no. 9, pp. 1602–1625, 2012. [Online]. Available: http://www.mdpi.com/1996-1944/5/9/1602

[115] G. S. Haase and J. W. McPherson, "Modeling of interconnect dielectric lifetime under stress conditions and new extrapolation methodologies for time-dependent dielectric breakdown," in *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, April 2007, pp. 390–398.

[116] F. Chen and M. Shinosky, "Addressing cu/low-k dielectric tddb-reliability challenges for advanced cmos technologies," *IEEE Transactions on Electron Devices*, vol. 56, no. 1, pp. 2–12, Jan 2009.

[117] J. C. K. Lam, M. Y. M. Huang, T. H. Ng *et al.*, "Evidence of ultra-low-k dielectric material degradation and nanostructure alteration of the cu/ultra-low-k interconnects in time-dependent dielectric breakdown failure," *Applied Physics Letters*, vol. 102, no. 2, Jan 2013.

[118] M. R. Baklanov and K. Maex, "Porous low dielectric constant materials for microelectronics," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 364, no. 1838, pp. 201–215, Jan 2006.

[119] W. Volksen, R. D. Miller, and G. Dubois, "Low dielectric constant materials," *Chemical Reviews*, vol. 110, no. 1, pp. 56–110, Dec 2010.

[120] D. Young and A. Christou, "Failure mechanism models for electromigration," *IEEE Transactions on Reliability*, vol. 43, no. 2, pp. 186–192, June 1994.

[121] A. S. Oates, "Electromigration failure distribution of contacts and vias as a function of stress conditions in submicron ic metallizations," in *Proceedings of International Reliability Physics Symposium*, April 1996, pp. 164–171.

[122] B. Li, C. Christiansen, D. Badami *et al.*, "Electromigration challenges for advanced on-chip cu interconnects," *Microelectronics Reliability*, vol. 54, no. 4, pp. 712–724, April 2014.

[123] M. H. Lin, S. C. Lee, and A. S. Oates, "Electromigration mechanisms in cu nanowires," in *2010 IEEE International Reliability Physics Symposium*, May 2010, pp. 705–711.

[124] B. Li, C. Christiansen, and R. Filippi, "Variability challenges to electromigration (em) lifetime projections," in *2014 IEEE International Reliability Physics Symposium*, June 2014, pp. 5A.4.1–5A.4.5.

[125] R. de Orio, H. Ceric, and S. Selberherr, "Physically based models of electromigration: From black's equation to modern tcad models," *Microelectronics Reliability*, vol. 50, no. 6, pp. 775–789, March 2010.

[126] Z. Lu, W. Huang, M. R. Stan *et al.*, "Interconnect lifetime prediction for reliability-aware systems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 15, no. 2, pp. 159–172, Feb 2007.

[127] C. M. Tan and A. Roy, "Electromigration in ulsi interconnects," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 52, no. 1–2, pp. 1–75, Oct 2007.

[128] I. A. Blech, "Electromigration in thin aluminum films on titanium nitride," *Journal of Applied Physics*, vol. 47, no. 4, pp. 1203–1208, Nov 1976.

[129] C. V. Thompson, "Using line-length effects to optimize circuit-level reliability," in *2008 15th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, July 2008, pp. 1–4.

[130] D. Rodopoulos, G. Psychou, M. M. Sabry *et al.*, "Classification framework for analysis and modeling of physically induced reliability violations," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 38:1–38:33, Feb. 2015. [Online]. Available: http://doi.acm.org/10.1145/2678276

[131] S. V. Kumar, K. H. Kim, and S. S. Sapatnekar, "Impact of nbti on sram read stability and design for reliability," in *7th International Symposium on Quality Electronic Design (ISQED'06)*, March 2006, pp. 6 pp.–218.

[132] K. Kang, H. Kufluoglu, K. Roy *et al.*, "Impact of negative-bias temperature instability in nanoscale sram array: Modeling and analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 10, pp. 1770–1781, Oct 2007.

[133] S. K. Krishnappa and H. Mahmoodi, "Comparative bti reliability analysis of sram cell designs in nano-scale cmos technology," in *2011 12th International Symposium on Quality Electronic Design*, March 2011, pp. 1–6.

[134] B. Cheng, A. R. Brown, and A. Asenov, "Impact of nbti/pbti on sram stability degradation," *IEEE Electron Device Letters*, vol. 32, no. 6, pp. 740–742, June 2011.

[135] A. Bansal, R. Rao, J.-J. Kim *et al.*, "Impacts of nbti and pbti on sram static/dynamic noise margins and cell failure probability," *Micrelectronic reliability journal*, vol. 49, no. 6, pp. 642–649, June 2009.

[136] J. C. Lin, A. S. Oates, H. C. Tseng *et al.*, "Prediction and control of nbti – induced sram vccmin drift," in *2006 International Electron Devices Meeting*, Dec 2006, pp. 1–4.

[137] J. C. Lin, A. S. Oates, and C. H. Yu, "Time dependent vccmin degradation of sram fabricated with high-k gate dielectrics," in *2007 IEEE International Reliability Physics Symposium Proceedings. 45th Annual*, April 2007, pp. 439–444.

[138] P. Weckx, B. Kaczer, P. J. Roussel *et al.*, "Impact of time-dependent variability on the yield and performance of 6t sram cells in an advanced hk/mg technology," in *2015 International Conference on IC Design Technology (ICICDT)*, June 2015, pp. 1–4.

[139] V. P. H. Hu, M. L. Fan, C. Y. Hsieh *et al.*, "Finfet sram cell optimization considering temporal variability due to nbti/pbti, surface orientation and various gate dielectrics," *IEEE Transactions on Electron Devices*, vol. 58, no. 3, pp. 805–811, March 2011.

[140] S. Khan, I. Agbo, S. Hamdioui *et al.*, "Bias temperature instability analysis of finfet based sram cells," in *Proceedings of the Conference on Design, Automation & Test in Europe*, ser. DATE '14. 3001 Leuven, Belgium, Belgium: European Design and Automation Association, 2014, pp. 31:1–31:6. [Online]. Available: http://dl.acm.org/citation.cfm?id=2616606.2616644

[141] S. Khan, M. Taouil, S. Hamdioui *et al.*, "Impact of partial resistive defects and bias temperature instability on sram decoder reliablity," in *2013 8th IEEE Design and Test Symposium*, Dec 2013, pp. 1–6.

[142] R. Menchaca and H. Mahmoodi, "Impact of transistor aging effects on sense amplifier reliability in nano-scale cmos," in *Thirteenth International Symposium on Quality Electronic Design (ISQED)*, March 2012, pp. 342–346.

[143] J. Kinseher, L. Heiß, and I. Polian, "Analyzing the effects of peripheral circuit aging of embedded sram architectures," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2017*, March 2017, pp. 852–857.

[144] D. Kraak, I. Agbo, M. Taouil *et al.*, "Degradation analysis of high performance 14nm finfet sram," in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2018, pp. 201–206.

[145] N. Khoshavi, R. A. Ashraf, R. F. DeMara *et al.*, "Contemporary cmos aging mitigation techniques: Survey, taxonomy, and methods," *Integration, the VLSI journal*, vol. 59, pp. 10–22, June 2017.

[146] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and minimization of pmos nbti effect for robust nanometer design," in *2006 43rd ACM/IEEE Design Automation Conference*, July 2006, pp. 1047–1052.

[147] X. Yang and K. Saluja, "Combating nbti degradation via gate sizing," in *8th International Symposium on Quality Electronic Design (ISQED'07)*, March 2007, pp. 47–52.

[148] J. Chen, S. Wang, and M. Tehranipoor, "Efficient selection and analysis of critical-reliability paths and gates," in *Proceedings of the Great Lakes Symposium on VLSI*, ser. GLSVLSI '12. New York, NY, USA: ACM, 2012, pp. 45–50.

[149] S. Kothawade, D. M. Ancajas, K. Chakraborty *et al.*, "Mitigating nbti in the physical register file through stress prediction," in *2012 IEEE 30th International Conference on Computer Design (ICCD)*, Sept 2012, pp. 345–351.

[150] S. Khan and S. Hamdioui, "Modeling and mitigating nbti in nanoscale circuits," in *2011 IEEE 17th International On-Line Testing Symposium*, July 2011, pp. 1–6.

[151] K. Kang, H. Kufluoglu, M. A. Alam *et al.*, "Efficient transistor-level sizing technique under temporal performance degradation due to nbti," in *2006 International Conference on Computer Design*, Oct 2006, pp. 216–221.

[152] B. C. Paul, K. Kang, H. Kufluoglu *et al.*, "Temporal performance degradation under nbti: Estimation and design for improved reliability of nanoscale circuits," in *Proceedings of the Design Automation Test in Europe Conference*, vol. 1, March 2006, pp. 1–6.

[153] B. C. Paul, K. Kang, H. Kufluoglu *et al.*, "Negative bias temperature instability: Estimation and design for improved reliability of nanoscale circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 4, pp. 743–751, April 2007.

[154] L. Zhang and R. P. Dick, "Scheduled voltage scaling for increasing lifetime in the presence of nbti," in *Proceedings of the 2009 Asia and South Pacific Design Automation Conference*, ser. ASP-DAC '09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 492–497. [Online]. Available: http://dl.acm.org/citation.cfm?id=1509633.1509750

[155] L. Lai, V. Chandra, R. Aitken *et al.*, "Bti-gater: An aging-resilient clock gating methodology," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, no. 2, pp. 180–189, June 2014.

[156] S. Kiamehr, F. Firouzi, M. Ebrahimi *et al.*, "Aging-aware standard cell library design," in *2014 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2014, pp. 1–4.

[157] S. Kiamehr, M. Ebrahimi, F. Firouzi *et al.*, "Extending standard cell library for aging mitigation," *IET Computers Digital Techniques*, vol. 9, no. 4, pp. 206–212, 2015.

[158] M. Ebrahimi, F. Oboril, S. Kiamehr *et al.*, "Aging-aware logic synthesis," in *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2013, pp. 61–68.

[159] F. Oboril and M. B. Tahoori, "Aging-aware design of microprocessor instruction pipelines," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 5, pp. 704–716, May 2014.

[160] S. Wang, J. Chen, and M. Tehranipoor, "Representative critical reliability paths for low-cost and accurate on-chip aging evaluation," in *2012 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2012, pp. 736–741.

[161] E. Mintarno, J. Skaf, R. Zheng *et al.*, "Optimized self-tuning for circuit aging," in *2010 Design, Automation Test in Europe Conference Exhibition (DATE 2010)*, March 2010, pp. 586–591.

[162] E. Mintarno, J. Skaf, R. Zheng *et al.*, "Self-tuning for maximized lifetime energy-efficiency in the presence of circuit aging," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 5, pp. 760–773, May 2011.

[163] F. Oboril and M. B. Tahoori, "Reducing wearout in embedded processors using proactive fine-grain dynamic runtime adaptation," in *2012 17th IEEE European Test Symposium (ETS)*, May 2012, pp. 1–6.

[164] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in cmos circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 19, no. 4, pp. 603–614, April 2011.

[165] S. Gupta and S. S. Sapatnekar, "Employing circadian rhythms to enhance power and reliability," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 18, no. 3, pp. 38:1–38:23, Jul. 2013.

[166] N. Khoshavi, R. A. Ashraf, and R. F. DeMara, "Applicability of power-gating strategies for aging mitigation of cmos logic paths," in *2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS)*, Aug 2014, pp. 929–932.

[167] X. Chen, Y. Wang, Y. Cao *et al.*, "Assessment of circuit optimization techniques under nbti," *IEEE Design Test*, vol. 30, no. 6, pp. 40–49, Dec 2013.

[168] A. Calimera, E. Macii, and M. Poncino, "Nbti-aware clustered power gating," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 16, no. 1, pp. 3:1–3:25, Nov. 2010.

[169] F. Oboril and M. B. Tahoori, "Extratime: Modeling and analysis of wearout due to transistor aging at microarchitecture-level," in *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)*, June 2012, pp. 1–12.

[170] F. Oboril, F. Firouzi, S. Kiamehr *et al.*, "Negative bias temperature instability-aware instruction scheduling: A cross-layer approach," *J. Low Power Electronics*, vol. 9, no. 4, pp. 389–402, 2013. [Online]. Available: https://doi.org/10.1166/jolpe.2013.1284

[171] F. Firouzi, S. Kiamehr, and M. B. Tahoori, "Power-aware minimum nbti vector selection using a linear programming approach," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 1, pp. 100–110, Jan 2013.

[172] U. R. Karpuzcu, B. Greskamp, and J. Torrellas, "The bubblewrap many-core: Popping cores for sequential acceleration," in *2009 42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Dec 2009, pp. 447–458.

[173] J. Srinivasan, S. V. Adve, P. Bose *et al.*, "Exploiting structural duplication for lifetime reliability enhancement," in *32nd International Symposium on Computer Architecture (ISCA'05)*, June 2005, pp. 520–531.

[174] R. A. Ashraf, N. Khoshavi, A. Alzahrani *et al.*, "Area-energy tradeoffs of logic wear-leveling for bti-induced aging," in *Proceedings of the ACM International Conference on Computing Frontiers*, ser. CF '16, 2016, pp. 37–44. [Online]. Available: http://doi.acm.org/10.1145/2903150.2903171

[175] G. Psychou, D. Rodopoulos, M. M. Sabry *et al.*, "Classification of resilience techniques against functional errors at higher abstraction layers of digital systems," *ACM Comput. Surv.*, vol. 50, no. 4, pp. 50:1–50:38, Oct. 2017. [Online]. Available: http://doi.acm.org/10.1145/3092699

[176] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Comparative analysis of rd and atomistic trap-based bti models on sram sense amplifier," in *2015 10th International Conference on Design Technology of Integrated Systems in Nanoscale Era (DTIS)*, April 2015, pp. 1–6.

[177] I. Agbo, S. Khan, and S. Hamdioui, "Bti impact on sram sense amplifier," in *2013 8th IEEE Design and Test Symposium*, Dec 2013, pp. 1–6.

[178] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Integral impact of bti and voltage temperature variation on sram sense amplifier," in *2015 IEEE 33rd VLSI Test Symposium (VTS)*, April 2015, pp. 1–6.

[179] I. Agbo, M. Taouil, D. Kraak *et al.*, "Integral impact of bti, pvt variation, and workload on sram sense amplifier," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 4, pp. 1444–1454, April 2017.

[180] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Quantification of sense amplifier offset voltage degradation due to zero-and run-time variability," in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2016, pp. 725–730.

[181] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Comparative bti analysis for various sense amplifier designs," in *2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS)*, April 2016, pp. 1–6.

[182] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Read path degradation analysis in sram," in *2016 21st IEEE European Test Symposium (ETS)*, May 2016, pp. 1–2.

[183] I. Agbo, M. Taouil, S. Hamdioui *et al.*, "Bti analysis of sram write driver," in *2015 10th International Design Test Symposium (IDT)*, Dec 2015, pp. 100–105.

List of Publications

International Journals

1. **I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, H. Kükner, P. Weckx, P. Raghavan, F. Catthoor, *Integral Impact of BTI, PVT Variation, and Workload on SRAM Sense Amplifier*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 1444–1454, volume 25, issue 4, April 2017.

2. **I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, *Sense Amplifier Offset Voltage Analysis for both Time-zero and Time-dependent Variability*, ACM Transactions on Design Automation on Electronic Systems (TODAES), major revision.

3. **I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Impact and Mitigation of SRAM Read Path Aging*, Microelectronics Reliability Journal, accepted for publication.

4. D. Kraak, M. Taouil, **I. O. Agbo**, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, *Impact and Mitigation of Sense Amplifier Aging Degradation Using Realistic Workloads*, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 3464-3472, volume 25, issue 12, Dec. 2017.

International Conferences

1. D. Kraak, **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor. *Degradation analysis of high performance 14nm FinFET SRAM*, Design, Automation Test in Europe Conference Exhibition (DATE), 2018, pp. 201-206, Mar. 2018, Dresden, Germany.

2. D. Kraak, **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Mitigation of sense amplifier degradation using input switching*, Design, Automation Test in Europe Conference Exhibition (DATE), 2017, pp. 858–863, Mar. 2017, Swisstech, Lausanne, Switzerland.

3. **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, *Quantification of Sense Amplifier Offset Voltage Degradation due to Zero-and Run-Time Variability*, 2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), pp. 725–730, July 2016, Pittsburgh, USA. **Best Paper Award**

4. **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Read path degradation analysis in SRAM*, 2016 21th IEEE European Test Symposium (ETS), pp. 1–2, May 2016, Amsterdam, The Netherlands.

5. **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, *Comparative BTI analysis for various sense amplifier designs*, 2016 IEEE 19th International Symposium on Design and Diagnostics of Electronic Circuits Systems (DDECS), pp. 1–6, April 2016, Košice, Slovakia.

6. **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, *BTI analysis of SRAM write driver*, 2015 10th International Design Test Symposium (IDT), pp. 100–105, Dec. 2015, Dead Sea, Jordan.

7. **I. O. Agbo**, M. Taouil, S. Hamdioui, H. Kükner, P. Weckx, P. Raghavan, F. Catthoor, *Integral impact of BTI and voltage temperature variation on SRAM sense amplifier*, IEEE 33rd VLSI Test Symposium (VTS), pp. 1–6, April 2015, Napa, USA.

8. **I. O. Agbo**, M. Taouil, S. Hamdioui, S. Cosemans, P. Weckx, P. Raghavan, F. Catthoor, *Comparative analysis of RD and Atomistic trap-based BTI models on SRAM Sense Amplifier*, 2015 10th International Conference on Design Technology of Integrated Systems in Nanoscale Era (DTIS), pp. 1–6, April 2015, Naples, Italy. **Best Paper Award**

9. S. Khan, **I. O. Agbo**, S. Hamdioui, H. Kükner, B. Kaczer, P. Raghavan, F. Catthoor, *Bias Temperature Instability Analysis of FinFET Based SRAM Cells*, Proceedings of the Conference on Design, Automation & Test in Europe, DATE 2014, pp. 31:1–31:6, March 2014, Dresden, Germany.

10. **I. O. Agbo**, S. Khan, S. Hamdioui, *BTI impact on SRAM sense amplifier*, 2013 8th IEEE Design and Test Symposium, pp. 1–6, Dec. 2013, Marrakesh, Morocco.

INTERNATIONAL WORKSHOPS

1. D. Kraak, **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Sense Amplifier Offset Voltage Mitigation Under Presence of BTI*, RESCUE 2017 - Workshop on Reliability, Security and Quality, ETS 2017 Fringe Workshop, May 25-26, 2017, Limassol, Cyprus.

2. D. Kraak, **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, *On Mitigating Sense Amplifier Offset Voltage Degradation*, First IEEE International Workshop on Automotive Reliability & Test, Nov. 2016, Fort Worth, USA.

3. **I. O. Agbo**, M. Taouil, S. Hamdioui, H. Kükner, P. Weckx, S. Cosemans, P. Raghavan, F. Catthoor, W. Dehaene, *Comparative BTI Impact for SRAM Cell and Sense Amplifier Designs*, 2015 MEDIAN Finale - Workshop on Manufacturable and Dependable Multicore Architectures at Nanoscale, pp. 33–36, Nov. 2015, Tallinn, Estonia.

4. **I. O. Agbo**, M. Taouil, S. Hamdioui, H. Kükner, P. Weckx, P. Raghavan, F. Catthoor, *BTI Analysis for High Performance and Low power SRAM Sense Amplifier*, 2015 4th Workshop On Manufacturable and Dependable Multicore Architectures, pp. 1–4, Mar. 2015, Grenoble, France. **Best Paper Award**

5. **I. O. Agbo**, M. Taouil, S. Hamdioui, H. Kükner, P. Raghavan, F. Catthoor, *Impact of BTI on SRAM Sense Amplifier in the Presence of Temperature and Process Variation*, Joint MEDIAN–TRUDEVICE Open Forum, pp. 1, Sept. 2014, Amsterdam, The Netherlands.

POSTER PRESENTATIONS

1. **I. O. Agbo**, M. Taouil, D. Kraak, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Estimation of Sense Amplifier Offset Voltage Degradation due to Zero- and Run-time variability*, 2017 Biannual European - Latin American Summer School on Design, Test and Reliability (BELAS), May 8th - 10th 2017, Rotterdam, The Netherlands.

2. **I. O. Agbo**, M. Taouil, S. Hamdioui, P. Weckx, S. Cosemans, F. Catthoor, W. Dehaene, *Read Path Degradation Analysis in SRAM*, 2016 Biannual European - Latin American Summer School on Design, Test and Reliability (BELAS), May 30th - 1st June 2016, Torino, Italy. **Best Poster Presentation Award**

OTHER CONFERENCES

1. **I. O. Agbo**, S. Safiruddin, S. Cotofana, *Implementable building blocks for fluctuation based calculation in single electron tunneling technology*, 2009 9th IEEE Conference on Nanotechnology (IEEE-NANO), pp. 366–369, July 2009, Genoa, Italy.

CURRICULUM VITÆ

Innocent Okwudili Agbo was born in Amechi-Awkunanaw of Enugu-South local government area of Enugu state, Nigeria on 11th April, 1971. He attended secondary education at the Comprehensive secondary school Akpasha-Awkunanaw, Enugu where he obtained his SSCE certificate in 1990. Furthermore, He obtained his bacherlor's degree with second class honours upper division (2^1) from Computer Science and Engineering department of Enugu State university of science and technology (ESUT) in 1999. Thereafter, he obtained his master's diploma from Computer Engineering laboratory at Delft University of Technology, the Netherlands in 2010.

Then in April 2013, He joined the department of Quantum and Computer Engineering at the Faculty of Electrical Engineering, Mathematics and Computer Science at the same University under the supervision of Prof. dr. ir. Said Hamdioui to pursue his Ph.D degree. His research interests include the reliability analysis, design, monitoring, mitigation of memory systems, Computer arithmetic, nano electronics, logic design and computer architecture. He received the Best Paper Award at the IEEE Computer Society Annual Symposium on VLSI 2016, and the Best Paper Award at the International conference on Design and Test of Integrated Systems in the nano-era DTIS 2015.