Impact Analysis of NBTI/PBTI on SRAM $V_{\text{MIN}}$ and Design Techniques for Improved SRAM $V_{\text{MIN}}$

Tony Tae-Hyoung Kim and Zhi Hui Kong

Abstract—Negative bias temperature instability (NBTI) and positive bias temperature instability (PBTI) are critical circuit reliability issues in highly scaled CMOS technologies. In this paper, we analyze the impacts of NBTI and PBTI on SRAM $V_{\text{MIN}}$, and present a design solution for mitigating the impact of NBTI and PBTI on SRAM $V_{\text{MIN}}$. Two different types of SRAM $V_{\text{MIN}}$ (SNM-limited $V_{\text{MIN}}$ and time-limited $V_{\text{MIN}}$) are explained. Simulation results show that SNM-limited $V_{\text{MIN}}$ is more sensitive to NBTI while time-limited $V_{\text{MIN}}$ is more prone to suffer from PBTI effect. The proposed NBTI/PBTI-aware control of wordline pulse width and wordline voltage improves cell stability, and mitigates the $V_{\text{MIN}}$ degradation induced by NBTI/PBTI.

Index Terms—Negative bias temperature instability (NBTI), positive bias temperature instability (PBTI), SRAM $V_{\text{MIN}}$, SRAM, SRAM cell stability, pulsed wordline, wordline control

I. INTRODUCTION

Bias temperature instability (BTI) is one of the most challenging issues in circuit reliability. Negative bias temperature instability (NBTI) for PMOS transistors has been widely recognized as a serious concern in highly scaled CMOS technology while positive bias temperature instability (PBTI) for NMOS transistors also becomes prominent in high-k metal gate technology [1, 2]. The degradation of the device threshold voltage caused by NBTI and PBTI affects various circuit parameters, such as performance, leakage, and noise margin. When compared to logic circuits in an SRAM architecture, SRAM cells require more careful design considerations due to the smaller margin in cell stability, write ability, bitline swing, and timing. In addition to these issues, the NBTI and PBTI effects will further exacerbate the margin and can eventually lead to circuit failures. Therefore, robust SRAM design providing reliable operation under NBTI and PBTI reliability mechanisms is highly warranted.

A number of works have dealt with the impact of NBTI and PBTI on SRAM operation [3-6]. The threshold voltage degradation due to NBTI and PBTI decreases cell stability, which will become even more serious when combined with device mismatches [5, 6]. Li et al. analyzed the impact of NBTI on SRAM cell operation and claimed that the NBTI in an SRAM cell has a stronger impact on read and write speed than on static noise margin (SNM) particularly when the clock frequency is high [7]. The read speed degradation of an SRAM cell is more sensitive to PBTI, whereas the write margin is improved by NBTI and PBTI. Consequently, the minimum operating voltage, $V_{\text{MIN}}$ of SRAMs has been demonstrated to increase over time to compensate the degraded SNM [8].

Recently, ultra dynamic voltage scaling (UDVS) has attracted unprecedented interest since energy efficiency is widely recognized as the topmost design criterion in numerous applications such as micro-processors, portable electronics, wireless sensor node, etc., where $V_{\text{MIN}}$ is a major design concern for power and energy
saving during low-performance modes [9, 10]. It has been corroborated that minimum energy consumption can be achieved in the sub-threshold or near-threshold region where current drivability is an exponential function of device threshold voltage and node voltage. Therefore, threshold voltage degradations due to NBTI and PBTI affect \( V_{MIN} \) significantly.

In this paper, we analyze the impact of NBTI and PBTI on SRAM \( V_{MIN} \) and present an NBTI/PBTI-aware wordline control technique for improving \( V_{MIN} \). SNM-limited \( V_{MIN} \) can be improved through careful control of the wordline pulse width while additional delay for enabling sense amplifiers is required for NBTI and PBTI when \( V_{MIN} \) is determined by the performance requirement rather than the stability requirement. The remainder of the paper is organized as follows. Section II presents an overview of various aging mechanisms such as NBTI, PBTI, HCI, and TDDB. In Section III, we analyze the impacts of NBTI and PBTI on SRAM operations. Section IV discusses the NBTI/PBTI-aware wordline control for SRAM \( V_{MIN} \) improvement. Finally, the paper concludes in Section V.

II. OVERVIEW OF AGING MECHANISMS

NBTI [11-14] in PMOS transistors is recognized as the major reliability concern in advanced CMOS process technologies. This mechanism is characterized by a positive shift in the absolute value of the PMOS \( V_{th} \). It occurs when a device is biased in strong inversion without large \( V_{DS} \). The increase in \( V_{th} \) is attributed to hole trapping at the gate dielectric interface or into the gate dielectric. Particularly, the broken Si-H bonds at the gate dielectric interface generate positively charged traps for holes in the inversion layer as shown in Fig. 1(a) [15].

One interesting feature of NBTI is “recovery”. When a stressed device is turned off, it instantly enters the recovery phase where trapped holes in the gate dielectric and at the gate dielectric interface are released. In addition, the broken Si-H bonds at the gate dielectric interface are annealed, thereby reducing the absolute value of the device \( V_{th} \) (Fig. 1(b)). PBTI in NMOS transistors was not critical in silicon dioxide dielectrics. However, as the high-k dielectric stacks started to be employed from the 32 nm technology node, its contribution to the aging also becomes significant [1, 2].

Fig. 1(c) and (d) explains the degradation and the recovery of the NMOS \( V_{th} \). In contrast to NBTI, PBTI is generally understood to occur due to the electron trapping within the gate dielectric or at the gate dielectric interface. Compared to NBTI, the aging mechanism of PBTI has not been fully understood yet since high-k dielectrics are relatively new, and various material combinations are used along with different deposition techniques. Even though various models have been proposed to explain PBTI effect in high-k gate stacks, no model is fully supported by researchers.

Hot carrier injection (HCI) degradation appears when a device is biased in strong inversion with large \( V_{DS} \) [16]. Hot carriers with high kinetic energy collide with other atoms and carriers, and lead to secondary carriers generated through impact ionization (Fig. 1(e)). Some of the carriers can be injected into the gate dielectric, which creates traps at the gate dielectric interface and the dielectric bulk. The traps capture carriers, hence the device \( V_{th} \) increases. HCI is more prominent in NMOS devices, which face a smaller potential barrier than holes.
at the gate oxide interface.

Finally, time dependent dielectric breakdown (TDDB) occurs when the voltage across the gate stack creates traps within the dielectric \[17, 18\]. The traps can link together and form a conductive path through the gate dielectric, which is also known as oxide breakdown (Fig. 1(f)). TDDB is a significant reliability issue as the gate dielectric thicknesses are following an aggressive scaling trend while the voltage across the gate dielectric stacks is not scaled at the same rate. A smaller number of traps can form a conducting path due to the thin dielectric in the nanometer range. The adoption of high-k dielectrics improves TDDB, but TDDB remains a critical aging mechanism in those relatively new high-k materials.

Fig. 2 illustrates the aforementioned aging mechanisms of CMOS devices under standard digital logic operation [19]. As explained in the above paragraphs, NMOS devices suffer from HCI and PBTI, and PMOS devices undergo HCI and NBTI. TDDB is common in both device types. In digital logic gates, HCI occurs during switching period where short circuit current and dynamic current flow while NBTI and PBTI appear after switching. It is worthy to note that no time period is free of aging.

In the rest of this paper, we will mainly focus on NBTI and PBTI since they are more prominent than HCI in SRAMs where switching probability of cell data is substantially low.

III. IMPACTS OF NBTI AND PBTI ON SRAM OPERATIONS

Undesirable changes in transistor characteristics as a result of NBTI and PBTI are primary reliability concerns in nano-scale CMOS technologies. This is particularly true in SRAMs since the selection frequency of an SRAM cell is substantially low such that it is nearing a DC bias stress condition. A conventional 6T SRAM cell with the indication of NBTI and PBTI is illustrated in Fig. 3. The PMOS transistor whose gate is storing data ‘0’ is under NBTI stress. Similarly, PBTI stress is applied to the NMOS transistor whose gate is storing data ‘1’. An SRAM cell needs to be designed to meet the required criteria such as stability, read ability, write margin, etc. Therefore, the degradation of devices in the SRAM cell over time has to be meticulously considered from the design stage to ensure its reliability in the long run. In this section, we will explain the impacts of NBTI and PBTI on SRAM read/write operation.

1. Read Operation

A read operation starts by enabling the wordline of the selected row and the pre-charged bitline pair is conditionally discharged through the access transistors and pull-down transistors. Two critical design parameters to be considered are stability and read delay. To limit the stability degradation caused by the disturbing current from the bitline to the cell node storing data ‘0’, the access transistors are designed to be weaker than the pull-down transistors. The read delay is determined by the series resistance formed by the access transistors and the pull-down transistors. It is evident that the PBTI of

![Fig. 2. Illustration of BTI mechanism (a) NBTI stress, (b) NBTI recovery, (c) PBTI stress, (d) PBTI recovery [19].](image)

![Fig. 3. Schematic of a 6T SRAM cell with NBTI and PBTI illustration [20].](image)
the pull-down transistors will exacerbate both the stability and the read delay, whereas the NBTI in the pull-up transistors will mainly decrease the stability.

2. Write Operation

During a write operation, write data is loaded into a bitline pair and the wordline of the selected row is enabled. The current strength of the access transistors should be greater than that of the pull-up transistors for pulling down the node of ‘1’ below the trip point of the other inverter. Once this occurs, the positive feedback of the cross-coupled inverters shifts the cell nodes from weak ‘0’ and ‘1’ to strong ‘0’ and ‘1’, and completes the write operation. This feature is characterized by write margin, which is defined as the highest bitline voltage of data ‘0’ that can successfully flip the stored data. The write margin is determined by the strength ratio of the access transistors to the pull-up transistors. Therefore, the NBTI in the pull-up transistors improves the write margin due to the weakened device strength.

IV. ANALYSIS OF SRAM $V_{MIN}$ DEGRADATION DUE TO NBTI AND PBTI

In this section, we will analyze the impact of NBTI and PBTI on SRAM $V_{MIN}$ using Predictive Technology Model (PTM) high-k 32 nm models. For the threshold voltage degradation due to NBTI and PBTI over time, the measured data in [1] and [2] is adopted for HSPICE simulation, which is also described in Fig. 4. Since NBTI and PBTI improve the write margin as mentioned in the previous section, we will mainly focus on the $V_{MIN}$ limited by read operation. In this work, we propose two different types of $V_{MIN}$: SNM-limited $V_{MIN}$ and Time-limited $V_{MIN}$. The remainder of this section will analyze the two different types of $V_{MIN}$, and explore the impact of NBTI and PBTI on the respective SRAM $V_{MIN}$.

1. SNM-limited $V_{MIN}$

In UDVVS systems, supply voltage is scaled aggressively to reduce power and energy consumption, and improve energy efficiency. If the energy efficiency is the topmost design constraint, sub-threshold operation can provide the minimum energy solution with substantially relaxed requirements in performance. In the sub-threshold region, $V_{MIN}$ is limited by the cell stability in read operation, which we define as SNM-limited $V_{MIN}$. Fig. 5 shows the SNM-limited $V_{MIN}$ degradation due to NBTI and PBTI respectively. It can be seen that NBTI has a stronger impact on the SNM-limited $V_{MIN}$ than PBTI. This can be explained based on the fact that the trip point of the inverters is formed at a lower voltage level due to the stronger pull-down transistors than the pull-up transistors. Thus the NBTI in the pull-up transistors affects the trip point more than the PBTI in the pull-down transistors. The combined effect of NBTI and PBTI (shown in Fig. 4) on the SNM-limited $V_{MIN}$ is demonstrated in Fig. 6. Over the stress time of $10^5$ seconds, SNM-limited $V_{MIN}$ degradation of 180 mV is obtained. For reliable SRAM operations over this period of time, $V_{MIN}$ degradation has to be continuously monitored and proper supply voltage has to be generated to compensate the reduced SNM.

![Fig. 4. Measured threshold voltage degradation due to NBTI and PBTI over stress time [1-2].](image)

![Fig. 5. Impact of NBTI and PBTI on SNM-limited $V_{MIN}$ respectively [20].](image)
Fig. 6. Degradation of SNM-limited $V_{\text{MIN}}$ over stress time. The NBTI and PBTI presented in Fig. 2 are used in this simulation [20].

![Graph showing SNM-limited $V_{\text{MIN}}$ degradation over stress time]

Fig. 7. (a) Impact of NBTI on read delay, (b) Impact of PBTI on read delay [20].

2. Time-limited $V_{\text{MIN}}$

Another circuit parameter to be considered as a $V_{\text{MIN}}$ limiting component is SRAM performance. This is particularly true when supply voltage is high enough to have positive SNM and SRAM performance is the primary design constraint. Fig. 7 demonstrates the effects of NBTI and PBTI on the read delay over various supply voltage levels. Note that the read delay is used as critical paths due to the slower performance than the write operation. The read delay is measured at the point where the bitline voltage of data ‘0’ is dropped by 10% of $V_{\text{DD}}$.

The NBTI in the pull-up PMOS transistors affects the data-flip process due to the slower performance than the write operation of the cross-coupled inverters. It is also shown that the speed of the data-flip becomes higher [6].

Fig. 8 summarizes the simulation results of the time to data-flip affected by the NBTI and PBTI in an SRAM cell with varying supply voltage. As expected, the time to data-flip decreases as the strength of the NBTI and the PBTI increases at a given supply level. In addition, raising supply voltage increases the time to data-flip since the raised SNM slows down the positive feedback operation of the cross-coupled inverters. It is also shown that the time to data-flip is less sensitive to PBTI than NBTI, whose reason is the same as that explained in Fig.

V. NBTI/PBTI-aware Wordline Control for Improving SRAM $V_{\text{MIN}}$

In this section, we will discuss NBTI/PBTI-aware design techniques for improving $V_{\text{MIN}}$ limited by SNM.

1. Data-flip Time Dependency on NBTI/PBTI

Once the SNM of an SRAM cell becomes zero or negative, data-flip occurs through the disturbing current from bitlines to cell nodes. As the SNM becomes more negative, the speed of the data-flip becomes higher [6]. This is because the more negative SNM provides larger voltage disturbance into the cell nodes and the positive feedback of the cross-coupled inverters amplifies it faster. Fig. 8 summarizes the simulation results of the time to data-flip affected by the NBTI and PBTI in an SRAM cell with varying supply voltage. As expected, the time to data-flip decreases as the strength of the NBTI and the PBTI increases at a given supply level. In addition, raising supply voltage increases the time to data-flip since the raised SNM slows down the positive feedback operation of the cross-coupled inverters. It is also shown that the time to data-flip is less sensitive to PBTI than NBTI, whose reason is the same as that explained in Fig.
5. In this simulation, no data-flip was observed from the supply voltage of 0.3V with PBTI of up to 120 mV. From Fig. 8, it can be inferred that the data-flip can be avoided if we complete read operation before the data-flip happens. To achieve this, it is necessary to estimate the time period where we can sense the bitline voltage and disable the selected wordline after sensing. This time period, which is defined as a sensing window in this work, can be calculated by subtracting the read delay from the time to data-flip. The following sub-section will explain this with more details.

2. Pulsed Wordline for Mitigating NBTI/PBTI Effect

A comprehensive NBTI and PBTI simulation result describing the time to data-flip, the sensing window, and the bitline swing is presented in Fig. 9. Fig. 9(a) uses the supply voltage that is lower than the $V_{\text{MIN}}$ of the SRAM cell by 10 mV. This is to generate negative SNM and data-flip accordingly. The simulation result shows that all the parameters including time to data-flip, sensing window and bitline swing, decline as the amount of the stress increases. However, since the SRAM cell has positive sensing windows, read operations can be conducted successfully through a careful wordline and sense amplifier control. Fig. 9(b) shows the same simulation result where the supply voltage is updated by the SNM-limited $V_{\text{MIN}}$ at a given stress time point. Due to the raised supply voltage, all the parameters are improved compared to Fig. 9(a) providing larger timing margins and bitline voltage swing. Note that the time to data-flip and the sensing window drops dramatically from the stress time of $10^4$ seconds. This phenomenon occurs since the supply voltage is raised high enough to cause the SRAM to operate in the strong inversion region where the read speed and the data-flip speed are higher than those in the sub-threshold region.

The effect of the NBTI/PBTI-aware pulse width control on read operation and $V_{\text{MIN}}$ is illustrated by a
simulation example shown in Fig. 10. Fig 10(a) shows the waveforms of a wordline (WL), a bitline pair (BL and BLB), and internal cell nodes (Q and QB) without NBTI and PBTI. No device mismatch is also assumed in the simulation for simplicity. It can be seen that data flipping is not occurring since the SNM of the cell is positive. Fig. 10(b) demonstrates the read operation when the NBTI in the stressed pull-up PMOS transistor is assumed to be 120 mV. The NBTI of 120 mV makes the SNM negative and eventually leads to data-flip. A conventional method of compensating the degraded is to use boosted supply voltage. Fig. 10(c) shows that the data-flip caused by NBTI of 120 mV at 0.3 V can be avoided by using 0.4 V. However, this leads to larger power consumption and requires non-trivial circuit overheads of generating 0.4 V out of 0.3 V. In Fig. 10(b), it can be seen that the bitline pair has a positive voltage margin at the early stage of the read operation. Therefore, if the wordline pulse width can be accurately controlled through the tracking of the NBTI and PBTI, the data-flip can be prevented by reading data and disabling wordlines as quickly as possible. Accordingly, the \( V_{\text{MIN}} \) of an SRAM is improved. Fig. 10(d) shows the case where the wordline pulse width is controlled after tracking the amount of NBTI and PBTI. The pulsed wordline avoid the data-flip even though the SNM is negative.

3. Wordline Voltage Control for Mitigating NBTI/PBTI Effect

Wordline pulse width control improves SRAM \( V_{\text{MIN}} \) by finishing read operation prior to data-flip. However, the continuously increased amount of aging leads to faster data-flip hence narrowing the bitline sensing window (Fig. 9(a)). To overcome this issue, design techniques for enhancing cell stability without significant
performance degradation are highly demanded. Boosted wordline voltage has been proposed to improve the write margin of SRAM cells [21], which degrades cell stability further due to the strengthened disturbance. It is only implementable when the cell stability still meets the requirement with the boosted wordline voltage. With NBTI and PBTI causing the cell stability marginal or negative, the boosted wordline scheme cannot be employed. In this work, we propose an aging-aware wordline voltage lowering scheme to improve the cell stability without raising supply voltage. Fig. 11 demonstrates the effects of lowering the wordline voltage on the cell stability. Note that data-flips occur when the wordline voltage is 300 mV in both NBTI and PBTI examples. In addition, lowering the wordline voltage increases the time to data-flip or, in the best-case scenario, eliminates the data-flips by mitigating the disturbance.

Fig. 12 shows the dependency of the time to data-flip and the read delay on the wordline voltage scaling and NBTI. It can be seen that the read delay dependency on NBTI is negligible while the time to data-flip dependency on the wordline voltage and NBTI is substantial. However, the lowered wordline voltage deteriorates the strength of the access devices, leading to an increase in the read delay. The read delay is primarily determined by the combined strength of access NMOS devices and pull-down NMOS devices. Therefore, lowering the wordline voltage increases the read delay. Fig. 13 and 14 summarizes the relationship between the read delay, NBTI and PBTI, and the wordline voltage. Note that the read delay with the lowered wordline voltage is still smaller than the time to data-flip. This validates that the proposed wordline voltage lowering scheme is efficient in improving the cell stability after NBTI and PBTI stress.
Fig. 13. Effects of wordline voltage lowering and NBTI on the read delay. NBTI has insignificant impact on the read delay. The read delay is measured at 0.5 × VDD.

Fig. 14. Effects of wordline voltage lowering and PBTI on the read delay. PBTI increases the read delay. Note that read operation fails with the threshold voltage degradation of 120 mV at VWL=300 mV.

4. Circuit Implementation of the Proposed Wordline Control Schemes

In this section, we will briefly discuss the circuit implementation of the proposed techniques. Fig. 15 illustrates simplified diagrams of the proposed technique implementation. The amount of the disturbance is sensitive to both PVT variations and the amount of NBTI and PBTI, which requires sensing blocks. Since the stability of the SRAM cell is the parameter to be monitored and sensed, the sensor should be implemented using the SRAM cell. Multiple SRAM cells should be included in the sensing block for better tracking device variations. The sensing results need to be digitized for reliable control operation at low-voltage operation where analog circuits are extremely difficult to be designed. Then, the digital codes will control either the pulse width or the voltage level of the wordlines. The sensors will check the stability again using the updated wordline pulse width or the updated wordline voltage. Finally, this feedback loop will continuously adjust the pulse width or the voltage level until the stability checking is passed.

VI. CONCLUSIONS

Negative Bias Temperature Instability (NBTI) and Positive Bias Temperature Instability (PBTI) are challenging reliability concerns in nano-scale CMOS technologies. In this work, we have analyzed the impact of NBTI and PBTI on SRAM VMIN. In the strong inversion region, SRAM VMIN is mainly limited by read delay, which has strong sensitivity to PBTI. However, in the sub-threshold region, SNM is the primary limiting parameter. The SNM-limited VMIN is significantly affected by NBTI. Therefore, different approaches have to be taken for either compensating or improving the two different VMINs. The proposed wordline pulse width control improves the SNM-limited VMIN mitigating the impact of NBTI and PBTI on SRAM VMIN. The wordline pulse width control combined with the supply voltage update improves the time to data-flip, the sensing window, and the bitline swing, providing better functional reliability.

ACKNOWLEDGMENTS

This work was supported by NTU Start-up Grant and Academic Research Fund (AcRF) Tier1, Singapore. The grant numbers for the project are M4080139 and M4010354 (RG23/10).
REFERENCES


Tony Tae-Hyoung Kim received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, Korea, in 1999 and 2001, respectively. He received the Ph.D. degree in electrical and computer engineering from University of Minnesota, Minneapolis, MN, USA in 2009. From 2001 to 2005, he worked for Samsung Electronics where he performed research on the design of high-speed SRAM memories, clock generators, and IO interface circuits. In 2007 ~ 2009 summer, he was with IBM T. J. Watson Research Center and Broadcom Corporation where he performed research on isolated NBTI/PBTI measurement circuits and SRAM Mismatch measurement test structures, and battery backed memory design, respectively. In November 2009, he joined Nanyang Technological University, Singapore, where he is currently an assistant professor and a program director of VIRTUS: IC Design Centre of Excellence. He received the ISOCC 2011 Best Paper Award, 2008 AMD/CICC Student Scholarship Award, 2008 Departmental Research Fellowship from Univ. of Minnesota, 2008 DAC/ISSCC Student Design Contest Award, 2008, 2005, and 1999 Samsung Humantec Thesis Awards, and 2005 ETRI Journal Paper of the Year Award. He is an author/coauthor of +40 journal and conference papers and has 17 US and Korean patents. His research interests include low power and high performance digital, mixed-mode, and memory circuit design, ultra-low voltage sub-threshold circuit design for energy efficiency, variation and aging tolerant circuits and systems, and circuit techniques for 3D ICs.

Zhi Hui Kong received the B.Eng. (Hons) (Elect) degree from University of Technology, Malaysia in 2000 and Ph.D. (Elect. Eng.) degree from Nanyang Technological University (NTU), Singapore in 2006. Dr. Kong worked as a Teaching Fellow in 2007 and currently an Assistant Professor in the Division of Circuits and Systems, School of EEE, NTU. Dr. Kong has published extensively in internationally recognized journals and conferences, focusing primarily on ultra-low power and high-performance VLSI memory design and reliability solutions for memory subsystems for CMOS nano-scale integration. She was the recipient of the 2010 Best Paper Award and 2011 IEEE SSCS Seoul Chapter Award in International SoC Design Conference (ISOCC). Dr. Kong serves as a committee member and reviewer for various IEEE conferences and journals, including IEEE International Symposium of Circuits and Systems (ISCAS), IEEE Transactions on VLSI Systems (TVLSI), and IEEE Transactions on Circuits and Systems-I. She is the Principal Investigator (PI)/co-PI of several research projects including the S$50 million IC Design Centre of Excellence (VIRTUS) at NTU.