Reliability and Variability Analyses in SRAM Decoder

Seyab Khan Said Hamdioui
Computer Engineering Laboratory, Delft University of Technology,
Mekelweg 4, 2628 CD, Delft, The Netherlands
{M.S.K.Seyab, S.Hamdioui}@tudelft.nl

Abstract
Modern Static Random Access Memory (SRAM) systems are susceptible to reliability and variability issues. Examples of reliability and variability issues are Bias Temperature Instability (BTI) in the transistors and resistive/open interconnect defects, respectively. This paper analyzes the impacts of BTI and resistive defects independently as well as simultaneously on the SRAM decoder. First, the BTI analysis shows that the impact is strongly dependent on the selected wordline, transistor location, addressing scheme and can cause up to 14.27% additional delay to the address selection. Second, for resistive defects analysis, it shows that the pull-up and pull-down resistances can cause up to 23.65% and 16.95% additional delay, respectively. Finally, it shows that BTI impact accumulates with that of the resistances, resulting in an additional delay of up to 34.36% to the address selection of the decoder.

I. INTRODUCTION

Static Random Access Memory (SRAM) system is one of the driving forces in state-of-the-art semiconductor industry [1]. The memories suffer from variability and reliability issues [2]. Variability in the shrunk interconnects cause resistive/open defects. Reliability issues, such as, BTI -Negative Bias Temperature Instability (NBTI) in PMOS transistors and Positive Bias Temperature Instability (PBTI) in NMOS transistors- degrade transistors performance during operation [3–5].

From variability perspective, Nigh et al. in [6] suggested that newer technology generations will suffer from elusive variability defects that will escape the quality and test metrics but will cause timing failures during operation. Kundu in [7] argued that due to increasing use of copper in the interconnects, existing shorts and bridges turn into resistive/open defects causing additional delays. Needham reported in [8] that resistive defects are the main cause of field return for Intel microprocessor. On the other hand, BTI cause threshold voltage shift to MOS transistors during operation and consequently an additional gates/circuits delay [9]. In conclusion, resistive/open defects and BTI, both resulting in additional delays, are threats to SRAM system operation.

An SRAM system comprises cell arrays and peripheral circuits, such as address decoders, control circuits, read/write driver, and sense amplifier. Much have been published on resistive/open defects and BTI impact on the cell array [10–13]. For instance, Dilillom in [10] investigated resistive defect in SRAM cell. Nourivand et al. in [11] analyzed resistive defects in drowsy SRAM cell. Similarly, Kumar et al. in [12] analyzed NBTI impact on Static Noise Margin (SNM) and read stability of the SRAM cell. Kang et al. in [13] investigated BTI impact the on SRAM cell’s SNM, read and write stability, and leakage current. On the other hand, few authors have focused on defects analysis of the address decoders. For instance, Hamdioui et al. in [14] identified decoder delay faults due to inter and intra-gate resistive defects. Manoj in [15] proposed a test pattern to identify defects in the decoders.

Apart from the aforementioned contributions, the SRAM analysis still suffers from limitations, such as: (a) none of the published work has investigated BTI impacts on address decoders. As address decoders are activated in every Read/Write operations, they will suffer from BTI resulting in SRAM system performance degradation (b) no published work has considered the combined impact of BTI and resistive defects on SRAM decoder. Therefore, for reliable operation of the SRAM systems, it is essential to focus on the BTI induced degradation and resistive defects on the decoder circuits.

This paper presents an analysis of the degradation caused by BTI and resistive defects on the SRAM decoders. In this regard, the paper has the following contributions:

- Investigate BTI impact on the decoder for different addressing schemes. The different addressing schemes include linear, gray, and address-complement.
- Analyze the impact of resistive defects at different locations in the address selection circuits of the decoders.
- Investigate the combined impacts of BTI and resistive defects on the decoder.

The rest of the paper is organized as follows: Section II describes the overall SRAM and focuses on decoder. Thereafter, it presents BTI mechanism, model; and classifies the resistive defects in the decoder. Section III describes the experiments performed. Section IV analyzes the result of individual impacts of BTI and resistive defects. Section V presents the combined impact of resistive defects and BTI on the decoder. Finally, Section VI concludes the paper.

II. ANALYSIS FRAMEWORK

This section explains the role of decoder in an SRAM system. Thereafter, it describes BTI mechanism and its impact on the delay. Finally, it describes resistive defects in different locations of the decoder.
A. SRAM Decoder

A SRAM system consists of memory cell array, registers, control circuits and address decoders as shown in Fig. 1(a). The address decoder is the main focus of this paper and is responsible to activate a unique wordline in the memory based on the input address value. Two dimensional addressing schemes are used in the SRAM system, where the row decoder activates a single wordline and the column decoder selects the appropriate bitline.

The address selection circuit of a $5 \times 32$ decoder (5 input address bits and 32 output wordlines) is shown in Fig. 1(b). The circuit has a pull-up network that comprises five PMOS transistors connected in parallel and a pull-down network containing five NMOS transistors in series; both networks are connected to node $X$. A unique address sequence at the inputs $S_0, S_1, S_2, S_3$ and $S_4$ activates a unique wordline ($WL$). For instance, the decoder of the figure is activated for address $S_0 S_1 S_2 S_3 S_4 = 11111$ (i.e., selects WL$_{31}$).

B. Bias Temperature Instability

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

BTI Mechanism

BTI causes threshold voltage ($V_{th}$) increment to MOS transistors. The $V_{th}$ increment in a PMOS transistor that occurs under the 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.

Stress Phase: In the stress phase, the Silicon Hydrogen bonds ($\equiv$Si-H) break at Silicon-Oxide interface. The broken Silicon bonds ($\equiv$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:

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

where $N_0$, $k_f$, $k_r$, $k_H$, and $k_{H_2}$, represent initial bond density, $\equiv$Si-H breaking rate, $\equiv$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 $\equiv$Si-H breaking. However, the H atoms/molecules diffuse back towards the interface and anneal the $\equiv$Si- bonds. The number of interface traps that do not anneal by the approaching H atoms during the relaxation phase is given by [12]:

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

where $N_{IT}(t_o)$ is the number of interface traps at the start of the relaxation, $\xi$ is a relaxation coefficient with $\xi=0.5$ [12], $t_o$ is the duration of the previous stress phase and $t_r$ is the relaxation duration.

The $N_{IT}$ oppose the gate stress resulting in the threshold voltage increment ($\Delta V_{th}$). The relation between $N_{IT}$ and $\Delta V_{th}$ is given by [20]:

$$\Delta V_{th} = (1 + m) \cdot q \cdot N_{IT} / C_{ox} \cdot \chi \cdot \gamma,$$
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. \( \chi \) is a BTI coefficient with a value \( \chi=1 \) for NBTI and \( \chi=0.5 \) for PBTI [18]. Additionally, \( \gamma \) represents the stress duration with respect to the total input period (i.e., activity factor) of the transistor. The \( \gamma \) dependence of the \( \Delta V_{th} \) shows that transistors in a gate/circuit that have different stress and relaxation phases will suffer from different degradations.

**BTI Model**

BTI induced \( \Delta V_{th} \) of each individual MOS transistor has its contribution to the additional delay. A generalized formula that relates BTI induced \( \Delta V_{th} \) in a transistor to wordline/bitline delay is given by [19,20]:

\[
\Delta D = \frac{n.\Delta V_{th}}{(V_{gs} - V_{th})}
\]

(4)

where \( n \) is the velocity saturation index. Since NBTI causes \( \Delta V_{th} \) to PMOS transistor and PBTI causes \( \Delta V_{th} \) to NMOS transistor; therefore, analysis in this paper integrate the \( \Delta V_{th} \) due to NBTI/PBTI to the PMOS/NMOS transistors to analyze their combined impact.

**C. Resistive Defects**

The resistive defects in the decoder of Fig. 1(b) can be classified as followings.

- **Inside pull-up network:** The PMOS transistors between the supply and the final/intermediate output node constitute the pull-up network. \( R_{pu} \) is an example of a defect in the pull-up network of the decoders shown Fig. 1(b).

- **Inside pull-down network:** The NMOS transistors between the final/intermediate output node and the ground manifest the pull-down network. There are many possible resistive defect locations in the pull-down network of the decoder (e.g., between each pair of the NMOS transistors). For simplicity, \( R_{pd} \) is the only considered resistive defect in pull-down network of the decoder shown in Fig. 1(b) and is adjacent to the ground.

The resistive defects contribute to the delay of wordline activation and or de-activation delays in the following ways. First, the resistance limits the current that (dis-)charges node \( X \). Second, the voltage drop in the resistive defects can be high enough to prevent transitions at node \( X \).

**III. SIMULATION SETUP**

This section describes the simulation environment and the type of experiments that are conducted in the analysis.

**A. Experimental Environment**

A netlist of a \( 5 \times 32 \) address decoder (Fig. 1(b) shows the wordline selection for the last word only) has been synthesized using 32nm PTM transistor models [21] and simulated using HSPICE. For the BTI analysis, the impact is augmented to the transistors with Verilog-A modules. Each module generate voltage shift that depend on the activity factor of the transistor. Similarly, in the resistive defect analysis, resistors are inserted at the locations shown in Fig. 1(b). The measurements are performed, when the signals reach 50\% of the \( V_{dd} \).

**B. Performed Simulations**

The experiments performed in this paper are of three kinds. The first set of experiments focus on the impact of BTI; the second set focus on resistive defects; and the final set analyze their combined impact.

The BTI impact analysis has the following four experiments:

1) Threshold voltage increment in the reference transistors (\( M_{ref1} \)) and (\( M_{ref2} \)) of Fig. 1(b).

2) Impact of BTI on the wordline activation and de-activation delay. The analysis are performed for all the 32 wordlines.

3) Impact of the location of the BTI affected transistors on the delay increment.

4) Impact of the workload, this is analyzed by considering different addressing schemes in the decoder.

The analysis for resistive defects constitute the following experiments:

1) Impact of the resistive defect in the pull-up network on the wordline delay.

2) Impact of the resistive defect in the pull-down network on the wordline delay.

3) Impact of the resistive defect in the pull-down and the pull-down network on the wordline delay.

Finally, experiments are performed that combine (Experiment 2 from the first set and Experiment 3 from the second set) both BTI and resistive defects.

**IV. SIMULATION RESULTS**

This section presents and analyzes first two sets of experiments described in the previous section.
A. Impact of BTI

This section presents a comprehensive analysis of the BTI impact on decoders. Each of these experiments described in the previous section is detailed next.

BTI Impact on Threshold Voltage

PMOS/NMOS transistors are affected by NBTI/PBTI leading to increments in their threshold voltages. As an example case, we analyze the threshold increments due to NBTI on the PMOS transistor (M\textsubscript{ref1}) and due to PBTI on the NMOS transistors (M\textsubscript{ref4}) (see Fig. 1(b) only. The impacts on both transistors are depicted in Fig. 2(a).

Figure 2(a) shows how the $V_{\text{th}}$ increases over time. The $\Delta V_{\text{th}}$ in PMOS transistor (M\textsubscript{ref1}) approaches 70mV after $10^8$ s of operation. This corresponds to an increment of 14.25% in the threshold voltage to the NBTI-free case. Similarly, the $\Delta V_{\text{th}}$ in NMOS transistors (M\textsubscript{ref2}) due to PBTI approaches 52mV; a relative increase of 10.59% to the PBTI-free case.

BTI Impact on Wordline Delay

The $\Delta V_{\text{th}}$ in MOS transistors due to BTI affect the wordline activation and de-activation delays. Both NBTI and PBTI impact the delays uniquely, i.e., NBTI in PMOS transistors affect de-activation and PBTI in the NMOS transistors affect rising transitions. In addition, the access patterns (or workload) stress the transistors differently, resulting in different delay. We analyze the workload in more detail later. Here, we assume a linear-up addressing scheme (i.e., from address 0 to address 31) for the $5 \times 32$ decoder. The results are shown in Fig. 2(b); the induced delays in the wordline activation and deactivation approach 16% (due to PBTI) and 7.50% (due to NBTI) respectively.

The figure shows that BTI impact is more significant on the wordline activation. It can justified by analyzing Fig. 2(b). Normally, the NBTI induced delay is more significant than PBTI impact for an inverter [4]. However, as all the NMOS transistors of the pull-down network are connected in series, their impact accumulate, resulting in higher impact on the wordline activation. The figure also shows an oscillating trend in the impact as we observe from lower to higher wordlines. This is directly related to the unique address selection circuits for the wordline. For example, the address selection circuit in Fig. 2(b) activates wordline $WL\textsubscript{31}$. In contrast, the selection circuit for wordline $WL\textsubscript{0}$ has inverted address as inputs to the transistors. The oscillation is caused due different in the number of these inverters.

BTI Impact of Transistor Locations

The majority of published work assumes that $\Delta V_{\text{th}}$ in any transistor has uniform contribution to the delay increment [12,20]. However, in reality, contribution due to a given $\Delta V_{\text{th}}$ depends on the location of the degraded transistors. This location dependency of NBTI is analyzed for reference transistors M\textsubscript{ref1} and M\textsubscript{ref3}; for BPTI reference transistors M\textsubscript{ref2} and M\textsubscript{ref4} shown in the decoder of Fig. 1(b). For both cases, we assume the BTI in a single transistor at a time.

Fig. 3(a) shows the NBTI induced $\Delta V_{\text{th}}$ for the locations of M\textsubscript{ref1} and M\textsubscript{ref3}. The additional delay due to M\textsubscript{ref1} is 6.18% after $10^8$ s of operation, while the delay increment due to M\textsubscript{ref3} is of similar value (5.82%). Hence, for parallel connected transistors NBTI impact is nearly the same and the uniform approach in [12,20] is valid. In the right side of the figure, the results of PBTI impact in M\textsubscript{ref2} or M\textsubscript{ref4} are depicted (see Fig. 3(b)). The figure clearly shows the significance of the location of transistor that is affected by PBTI. For example, the induced $\Delta V_{\text{th}}$ in M\textsubscript{ref2} cause 2.01% additional delay, while the same shift in M\textsubscript{ref4} contribute 3.41% increment to the delay. The higher degradation in case of M\textsubscript{ref4} can be explained as follows. The increased $\Delta V_{\text{th}}$ at M\textsubscript{ref2} cause a much weaker drive strength at the source of transistor M\textsubscript{ref4}. Therefore, the uniform approach in [12,20] is not accurate enough for serially connected transistors.
BTI Impact of Various Workloads

The $\Delta V_{th}$ is determined by stress and relaxation durations of the transistors. Therefore, BTI impact is strongly dependent on the applied input patterns (henceforth known as workload). This workload dependence is analyzed by applying various addressing schemes to the decoder. Fig. 4 shows the obtained results of both wordline activation and wordline de-activation additional delays for the following addressing schemes: linear-up (0,1,2,3,...,30,31), linear-down (31,30,29,28,...,1,0), gray-up (0,1,3,2,...,17,16), gray-down (16,17,19,18,...,1,0), address complement-up (0,31,1,30,...,15,16) and address complement-down (15,16,14,17,....,31,0). The result reveals that BTI impact on wordline activation ranges between 6.85% for address complement-down addressing scheme and 8.17% for the gray-up addressing scheme. Similarly, BTI impact on the worldline de-activation ranges between 12.69% and 14.27% for the address complement-down and linear-up addressing schemes, respectively. Hence, it can be concluded that the addressing schemes can cause significant variation in the BTI induced delays.

B. Impact of resistive network

The resistive defects in the pull-up or pull-down and their combination affect wordline activation and de-activation delays as described below.

Inside Pull-up Network Impact

The resistive defect in the pull-up network of the decoder shown in Fig. 1(b) affect the rising transition at node X, consequently de-activation of the wordline.

Figure 5(a) shows the wordline transitions for various resistive defects: $R_{pu}$ between 0kΩ and 100kΩ. In general, it shows that the resistance have impact on both activation and de-activation transitions of the decoder. However, impact on the wordline de-activation is more significant. The larger the resistive defects, the worse the transition delay. For example, the falling transition time increases by 6000% in case $R_{pu}=100k\Omega$ as compared to the defect free case. The higher resistive defects are easy to catch at test time. However, critical are the lower resistance defects as they can lead to test escapes and may cause problems in the field.

Fig. 5(c) shows the impact of these low resistances present in the pull-up and pull-down networks. Here, we focus on resistances in the pull-up network only. For the pull-up network, it depicts the de-activation delay increment for $R_{pu}$ between
Fig. 5. (a) Impact of resistance in the pull-down network (b) Impact of resistance in the pull-up network (c) Impact of the variable resistances in the pull-up and the pull-down networks.

Fig. 6. Impact of pull-up and pull-down resistances.

0 and 5 kΩ. The delay increments shows a linear behavior with increasing pull-up resistance values. The de-activation delay increases approximately with a rate of 3.74%/kΩ (18.7%/5kΩ).

**Inside Pull-down Network Impact**

The resistive defect in the pull-down network of the decoder shown in Fig. 1(b) affects activation delay of the wordlines. Simulation results of the observation are depicted in Fig. 5(b). The ascending resistances has significant impact on the wordline activation. For example, at 100 kΩ the resistance causes even a transition failure, i.e., the pull-down network is not strong enough to drive node X to zero and therefore, the inverter fails to drive the wordline high. Fig. 5(c) shows the impact of the low lower resistance $R_{pd}$ between 0 and 5 kΩ. The delay increments shows a linear behavior with increasing pull-down resistance values. The figure shows that activation delay increases with a rate of 4.96%/kΩ (24.8%/5kΩ).

**Combined Pull-up and Pull-Down Network Impact**

The previous two subsections inserted the pull-up and pull-down resistances separately and showed that they have mutually exclusive impacts on the wordline delays. In the current analysis both $R_{pu}$ and $R_{pd}$ are simultaneously considered in the decoder of Fig. 1(b). In this case de-activation and activation transitions are both affected. The figure shows that $R_{pu}=R_{pd}=0kΩ$, the mean delay (activation and de-activation) increment is zero. Furthermore, it shows that the delay increases with the $R_{pu}$ or $R_{pd}$ or their combination. The additional delay approaches 22% of the nominal at $R_{pu}=R_{pd}=5kΩ$.

In conclusion, it can be argued that testing for resistive defects should focus on both transitions independently to get the overall delay. To spot the exact location of the defect, de-activation or activation transitions should be considered.

**V. BTI AND RESISTIVE DEFECT IMPACT ON SRAM DECODER**

This section focuses on the last set of experiment described in Section III to analyze the combined impact of BTI and the resistive defects. The analysis are carried out for linear-up addressing scheme in the decoder shown in Fig. 1(b). It should be noted the analysis initially consider BTI and resistance in the pull-up or pull-down network. Finally, it consider BTI and both resistances simultaneously.
Figure 7 shows the analysis results for BTI and resistive defects $R_{pu}$ or $R_{pd}$ (with the values similar as in Fig. 5(c)). The figure shows the impact of BTI and resistances have mutually exclusive and different impact variation rate on the wordline delays. For example, if we consider $R_{pu}=R_{pd}=0$, the delay increment in the wordline activation is only 7.50%, while the delay increment in the de-activation is 13%. The BTI induced higher increment in the de-activation is consistant with results presented in the previous section. The figure shows that increments either in the $R_{pd}$ or in $R_{pu}$ increase activation and de-activation wordline delay, respectively. However, the increment rate in the de-activation due to $R_{pu}$ increment is higher than caused by $R_{pd}$ on the activation. The analysis is consistant with the $R_{pu}$ and $R_{pd}$ impacts observed in the previous section. Due to different impacts of BTI and resistive impacts, at $R_{pu}=R_{pd}=5k\Omega$ the additional delays due to BTI and resistances reaches about 33%.

Finally, the combined impact of the resistive defects ($R_{pu}$ and $R_{pd}$) and BTI on wordline activation and de-activation are performed. Results of the analysis are shown in Fig. 8. The figure shows that in absence of resistive defects (i.e. $R_{pu}=R_{pd}=0k\Omega$ ), BTI induced mean delay (delay to both activation and de-activation delays) is 12.40%. Furthermore, the mean delay increases with elevation in the resistive defects. The impact of only $R_{pu}$ increment on the mean delay approach 21.79%, while that of only $R_{pd}$ approach 22.07%. However, when both resistance are considered simultaneously, the mean delay increment approaches 34.36%. The combined impact is quite significant and may cause timing failure in the address decoder and consequently, the SRAM system.

VI. CONCLUSION

The analysis presented in this paper for BIT and resistive defects clearly showed that their combination cannot be ignored; this will become even severe for the future smaller technologies. As the simulation results showed, depending on the resistor location and the workload, the additional delay can approach up to 34.36%. Therefore, it is extremely important to address the degradation of the decoder performance both at the design stage and during operation. At the design stage, the test technique should be developed to capture the smaller resistive defects in the decoder. However, if the defects escape the test, it should be mitigated during the operation. During operation, the BTI impact exacerbate the delay due to resistive defects. Therefore,
their combined impact should be mitigated by dynamic techniques such as frequency reduction, $V_{dd}$ adjustment and adoptive body biasing.

REFERENCES