Reliability Implications of Bias-Temperature Instability in Digital ICs

Renju Raju Thomas

Technische Universität München

renju.thomas@mytum.de

Abstract

This paper is a brief overview of reliability concerns due to negative bias temperature instability (NBTI) associated mainly with PMOS transistors when they are subjected to negative bias at elevated temperatures and is a summary of reference papers [1] [2] [3] [4]. This paper gives a detailed description of the NBTI mechanism, interface traps and their origin. A reaction diffusion framework for NBTI degradation and recovery process, the effects of NBTI as a major reliability concern in nanometer designs at device level, random logic circuits and SRAM cells are also addressed. Finally, reliability aware circuit design techniques to ensure optimum performance during the entire life time of a circuit are also presented.

---

1. Introduction

Now a days, circuits and systems are growing in speed and complexity and a successful IC design grows to a complex optimization problem with several considerations like silicon area, speed, testability, design effort and power dissipation. This traditional design methodology assumes that the electrical and physical properties of transistors are deterministic and hence predictable over the life time of the device. With the continuous silicon technology scaling to nm range for better circuit operation speed, integration density and power consumption, transistor properties are no longer deterministic. It is the temporal

reliability degradation due to bias-temperature instability which makes the transistor properties no longer deterministic.

Negative bias-temperature instability (NBTI) is a transistor-aging effect and is mainly associated with p channel transistors. When the transistor is biased in inversion, holes are attracted to the Si/SiO_2 interface. These holes under high temperature will cause an increase in the threshold voltage of the device, over its life time. NBTI also reduces the carrier mobility as a function of time and stress conditions.

NBTI has been known since 1966 but it became a major reliability concern in nanoscale CMOS technology only during the last few years. The main reasons [2] are

(a) Increase in gate electric fields to a range of few megavolts per centimeter, due to scaling the oxides to less than 10 angstroms for technology nodes below 32nm.
(b) Increase in chip operating temperatures.
(c) Introduction of dual poly-process that has allowed replacement of buried channel PMOS devices with surface channel PMOS devices.
(d) Introduction of nitrogen to reduce gate leakage and to inhibit boron penetration in thin oxide.

The increase in threshold voltage, decrease in mobility, drain current and transconductance are the degradation effects which can result in a failure to meet the timing specifications. The life time and maximum frequency of operation of the device is reduced and if the system is a critical embedded system, it can raise severe safety concerns too. It is in this modern context that we need to understand the origin of NBTI and how to control it to ensure optimum performance during the entire life time of a circuit.

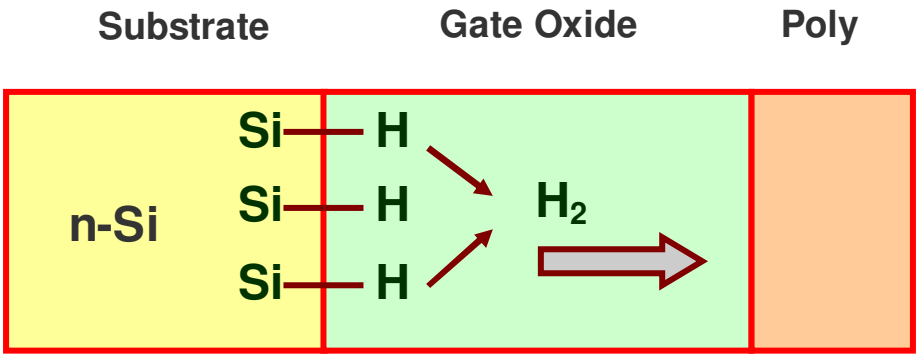
NBTI is still a challenging topic that is not yet fully understood and experimental results shows about 25-30% increase in V_{TH} over a period of 10 years [4]. The results show a wide variation with the circuit topology and operating conditions. The number of PMOS transistors in inversion, time for which they are stressed and the fact whether they are connected in series or parallel in a circuit are key factors which affect the degradation strength. The time delay between the application of stress and measurement of degradation is also quite important as it can produce a wide variation in the results.

The rest of the paper is organized as follows. In section 2, the mechanism of NBTI is discussed using the Reaction-Diffusion framework. In section 3 Reaction-Diffusion based V_{TH} model is introduced and how the NBTI degradation affects threshold voltage of a transistor is explained. In section 4 and 5 the effect of NBTI in random logic circuits and SRAM cells is discussed. Section 6 explains impact of statistical variation in NBI together with random dopant fluctuations in a circuit and section 7 describes the various reliability aware circuit design techniques. The paper is concluded in section 8.

2. NBTI Mechanism

One of the most prevalent models proposed to explain NBTI mechanism is the Reaction-Diffusion (R-D) model. Figure 1 shows the NBTI mechanism based on RD model [3]. This model explains NBTI as a result of continuous trap generation at Si/SiO_2 interface of the transistor. Silicon is tetravalent and shares its four valence electrons with other four Si atoms to complete the octet. But, at the $Si\text{-}SiO_2$ interface there can be dangling Si atoms. The hydrogen passivation process to remove these dangling silicon atoms result in Si-H bonds. These bonds at negative bias and elevated temperatures can easily break to produce a donor like state known as interfacial traps and neutral hydrogen atoms. This marks the reaction phase. The hydrogen

atoms can form H_2 and will diffuse away from the interface towards poly through SiO_2 or anneal existing traps. This is the diffusion phase. The rate of generation of these traps is accelerated by temperature and the duration of applied stress. The rate of diffusion of hydrogen through oxide and poly silicon also has an impact. When the stress is removed the hydrogen will diffuse back to the interface and this will cause a reduction in the degradation due to a reverse effect [3].



$$Si_3 \equiv SiH + H^+ \leftrightarrow Si_3 \equiv Si\bullet + H_2$$

Figure 1. NBTI mechanism in a PMOS transistor based on reaction diffusion (RD) model.

3. RD based V_{TH} model

The temporal V_{TH} model for NBTI degradation using RD framework is quite convincing. This analytical NBTI model based on reaction and diffusion can be used to determine the performance degradation in a single transistor. We can extend the model to evaluate the impact of NBTI degradation on both logic circuits and SRAM cells. NBTI introduces an increase in V_{TH} of the device due to the generation of traps at Si-SiO_2 interface [1]. In the past few decades numerous experiments were carried out with two types of stress- DC stress and AC stress. DC stress implies that the negative bias is continuously applied to the PMOS so that it is in inversion region throughout the period of analysis and AC stress means the polarity of the input signal changes during the period of analysis. The signal probability is a

crucial factor which affects the degradation and is defined as the fraction of time the input is high.

The change in V_{TH} can be accounted as follows [1] :

$$\Delta V_{TH}(t) \approx \frac{q \bullet N_{it}(t)}{C_{ox}}$$
$$\approx f_{AC}(S_p) \bullet K_{DC} \bullet t^{n}$$

where N_{it} is the density of interfacial traps, C_{ox} is the oxide capacitance and q is the charge of the electron. f_{AC} is the signal probability (S_p) dependent function which represents the AC dependency of the process. K_{DC} is a technology dependent constant that depends on temperature, V_{DD}, device geometry, oxide nitrogen concentration and various other factors. The time dependence is given by the factor t^n where n depends on the mode of measurement.

The increase in V_{TH} due to NBTI under DC stress follows a power law [1] with respect to time, t with an exponent, n ($V_{TH} \sim t^n$). The experimental dependency of the degradation process is given by t. On the fly measurements has indicated a value of n close to 1/6 and for measurements with delay indicated n~1/4. But, in reality transistors are subjected to a series of AC stresses with varying signal probabilities S_p. So with AC stress there are two phases of degradation - stress phase and recovery phase [1]. Figure 2 shows the stress and recovery phases in NBTI degradation. When PMOS is negatively biased it experiences a constant degradation and is the stress phase. But, once the stress is removed there is a tendency to revert back due to reverse annealing and is the recovery phase. The degradation under AC stress is weakly dependent on operating frequency and has a similar time exponent factor of 1/6. The absolute magnitude of degradation is scaled by a signal probability dependent function f_{AC}. Figure 3 shows the NBTI degradation under DC and AC stress with different signal probabilities. Due to the recovery phase, there is a tendency to revert back the degradation effects and so AC degradation is less than the DC degradation.

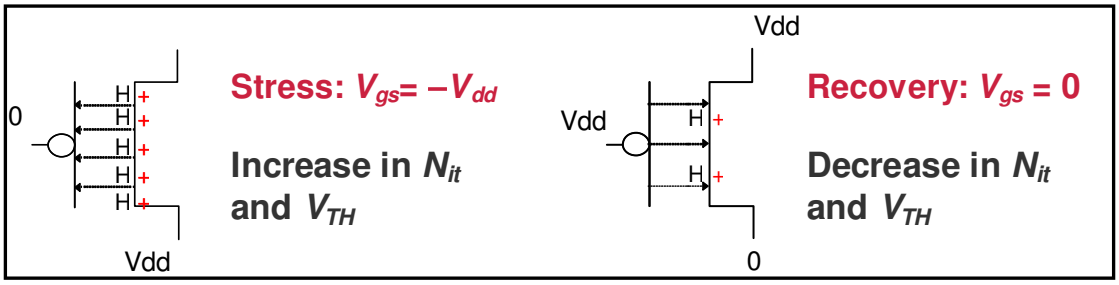


Figure 2. Stress and recovery phases of NBTI degradation under an AC stress.

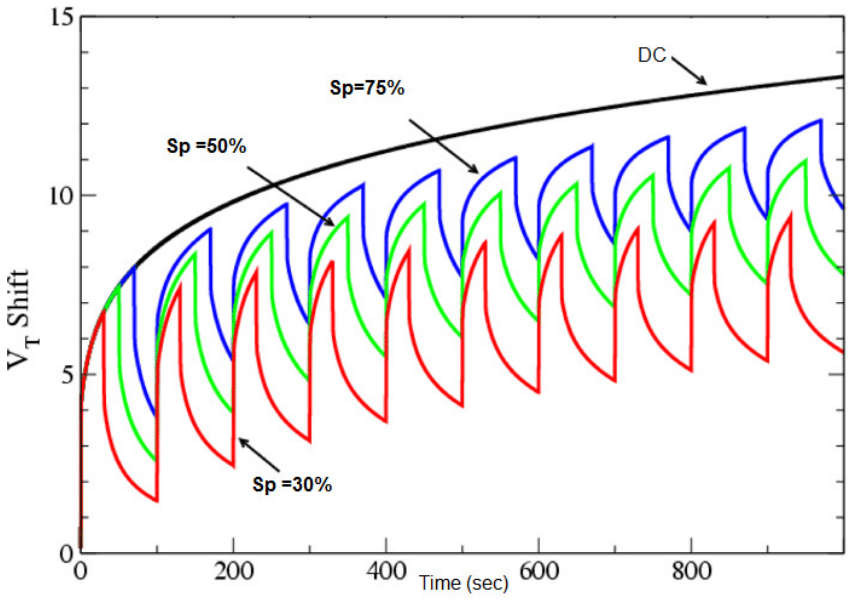


Figure 3. NBTI degradation under DC and AC stress with different signal probabilities.

4. NBTI in random logic circuits

NBTI causes performance degradation in the random logic circuits and can be estimated as delay degradation of the circuit. The increase in threshold voltage reduces the drive current and thus the individual gate delay is increased. The increase in delay directly causes the degradation in maximum frequency of operation and Figure 4 shows approximately 8.8 % degradation in maximum frequency of operation, f_{max} in 3 years (10^8 sec) at 125^0C [1]. But at 25^0C the degradation percentage is about 3.3% and this shows the importance of a realistic estimation of the degradation under the expected temperature profile. The degradation shows a power relation with respect to time with

exponent factor 1/6 which is identical to single device degradation case. The degradation in maximum operating frequency in a crucial real time embedded system can bring forward critical safety concerns. The final effect is reduction in reliability and life time of the device.

The performance degradation of a large circuit is relatively small compared to single device degradation due to the following reasons [1]:

- NBTI affects the rising transition in CMOS gates. In reality, a timing path has both rising and falling transitions and it will reduce the degradation effect. Delay degradation in a non-critical data path may not cause a timing failure.
- The sensitivity of different gates to NBTI is different. NOR gate with stacked PMOS has more degradation because the stacked transistors simultaneously affect transition.
- The signal probability at different nodes shows a wide variation and such an AC pattern brings down the effect of NBTI.

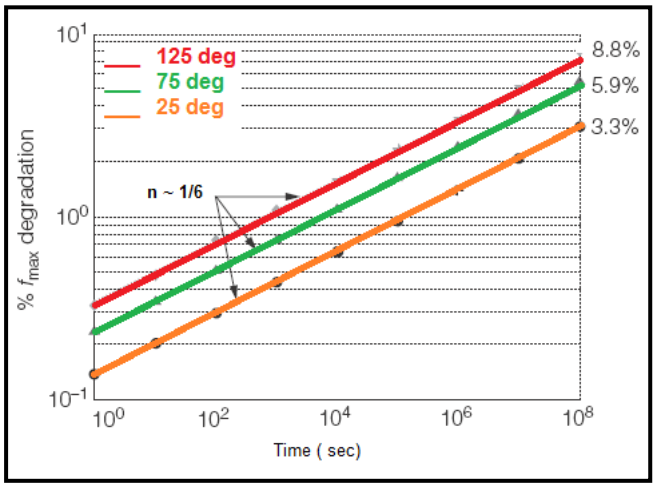


Figure 4. Maximum frequency of operation (f_{max}) degradation at different temperatures.

5. NBTI in 6T SRAM cell

Static RAM stores a bit in cross-coupled inverters. The output of one inverter is fed to the input of other and it forms a loop. This storage cell has two stable states to denote 0 and 1. The four transistors 1, 2, 3 and 4 form the cross-coupled inverters and the two additional

access transistors 5 and 6 serve to control the access to a storage cell during read and write. Read and Write are realized by using the control signals, Bit Line and Word Line. Figure 5 shows the 6T SRAM cell.

In the standby mode, WL is not asserted and the access transistors will disconnect the bit lines from cell. The cross-coupled inverters reinforce each other so that the bit is stored as long as supply is not turned off. In write cycle, the data is placed on the bit line and the inverse data on the inverse bit line. Then the access transistors are turned on by setting the word line to high. As the driver of the bit lines is much stronger it can assert the inverter transistors.

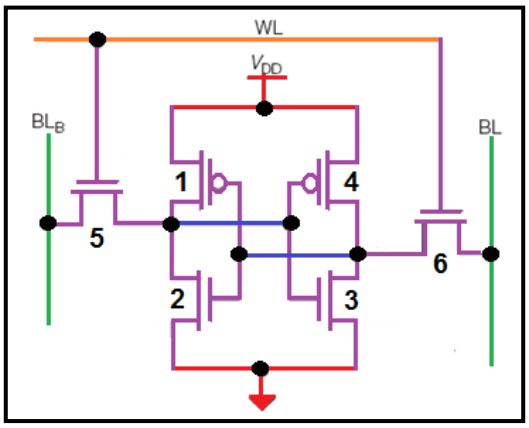


Figure 5. 6T SRAM cell

In read cycle, both bit lines are pre-charged to logical 1 and then WL is asserted which enables the access transistors. The values stored are transferred to the bit lines leaving BL at its pre-charged value and discharging inverse bit line through M2 and M5 to a logical 0. On the BL side, the transistors M4 and M6 pull the bit line toward V_{DD}, a logical 1.

The NBTI introduces degradation to PMOS transistors in the SRAM cell and the increased V_{TH} introduces mismatch among the six transistors which lead to failures [1]. If a cell stores a constant data for

a long time one PMOS transistor is more degraded which can further increase the mismatch. The read stability of SRAM cell is severely affected. Static Noise Margin (SNM) is a key figure of merit of SRAM cell and is defined as the side length of largest possible square between the two voltage transfer curves of the involved CMOS inverters when superimposed (Figure 6).

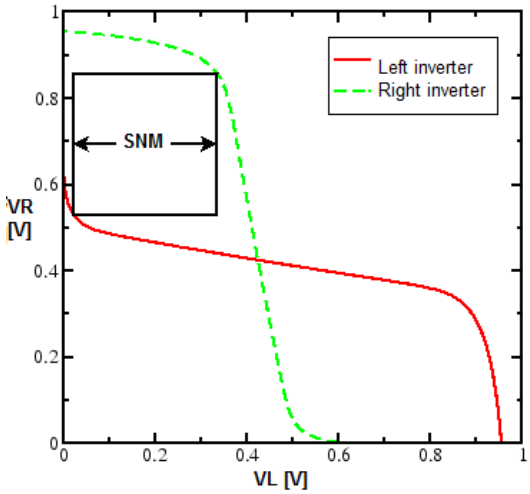


Figure 6. Static Noise Margin (SNM) of SRAM cell

When an external DC noise is larger than the SNM, the state of the SRAM cell can change and data is lost. The result of V_{TH} degradation is a degraded transfer curve and degraded SNM which in turn increases the read failure probability. The static noise margin of an SRAM cell can decrease by about 9% over a period of 3 years.

6. Statistical variation in NBTI

Transistor mismatch due to VTH variation usually originate from random dopant fluctuation (RDF) which is the variation in the number and placement of dopant atoms in channel. It is an intuitively pleasing result that NBTI combined with random dopant fluctuations has a greater effect. But, the breaking and repassivation of Si-H bonds which are randomly placed and varying in number experience stochastic fluctuations [1]. So a statistical NBTI model that considers this random nature of Si-H bonds is required and we consider

simultaneous RDF and NBTI induced VTH variation (σ_{RDF} and σ_{NBTI}) as follows [1]:

$$\sigma_{VTH} = \sqrt{\sigma^2_{RDF} + \sigma^2_{NBTI}(t)}$$

where σ_{VTH} represents total V_{TH} variation after time t.

The NBTI variations have a greater impact on scaled technology with more degradation in 22 nm node than in 32 nm node. In general, the degradation is more when we consider RDF & statistical NBTI together. In random logic circuits the standard deviation of the delay changes significantly because of this NBTI variation. But, in circuits with large logic depths, the variation cancelling effect makes it minimal. In SRAM cells, the read failure probability is significantly increased due to the degradation in SNM. The input signal probabilities for the PMOS transistors also have a greater impact on the degradation.

7. Reliability aware circuit design

Reliability aware circuit design has rapidly grown in prominence because NBTI can slow down the circuit throughout the life time and can possibly result in a timing failure in nanoscale technology designs. Delay guard-banding is the basic approach to compensate for the degradation and is an overdesign approach so that even after the degradation the circuit will function well. In random logic circuits, simple delay guard banding is a competent to complex reliability aware design techniques such as gate sizing and logic synthesis. However for memory arrays mismatch among transistors are crucial for read stability and hence more sophisticated circuit design techniques are used to reduce the impact of NBTI degradation over the life time of the device. The other design techniques to handle this degradation are described in the following sub sections [4].

7.1. Gate or Transistor sizing.

This is one of the first approaches of reliability aware circuit design where the size of the gate is changed to compensate for the degradation in setup timing margin with time. To overcome this, the designers put forward this over design approach by upsizing the transistors considering the signal probabilities. The guard band is selected as average delay degradation for 3 years. This approach will introduce an area overhead and to reduce the area overhead different sizing factors for NMOS and PMOS are considered. Figure 7 illustrates the increase in life time of a circuit when it is overdesigned to take care of the NBTI degradation.

The two sizing techniques are worst case sizing and optimal sizing [1].

- Worst case sizing.

  In worst case sizing approach, maximum degradation is considered and the circuit is designed so that it can compensate for future degradation.

- Optimal sizing.

  In the optimal sizing approach, the actual estimate of signal probabilities at all nodes is considered and the circuit is designed so that it can compensate for future degradation. It has less area overhead than worst case sizing.

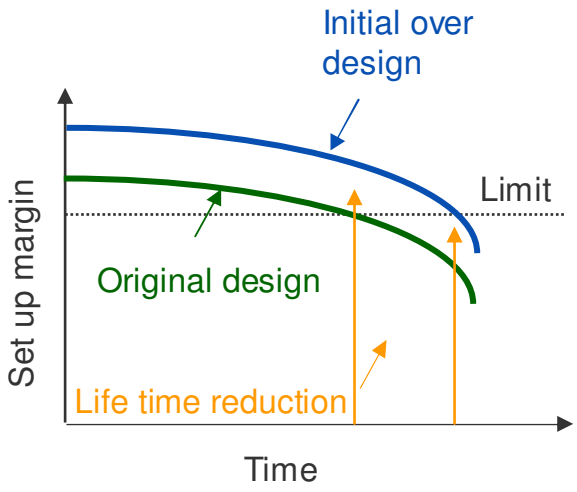


Figure 7. Reliability aware gate sizing to increase the life time of a circuit

7.2. Technology mapping and Logic synthesis.

An alternative approach is to consider the NBTI at technology mapping stage of logic synthesis. This technique considers the fact that different standard cells have different NBTI sensitivity with respect to their input signal probabilities. So the standard cell library is re-characterized with additional signal probability dependency. The two design approaches are worst case NBTI synthesis and signal probability based synthesis [4].

- Worst case NBTI synthesis.

  Technology mapping is done using the timing library with NBTI degradation effects. The worst case delay for gates computed for 10 years of continuous stress is used.

- Signal probability based synthesis.

  Technology mapping is done considering the signal probability at each node so that the gate with least area overhead to meet the timing specification is selected. The delay of the gates in the library as a function of signal probability is pre-calculated.

7.3. Self correction using on-chip sensor circuits.

Both the sizing and synthesis methods require an accurate estimation of the NBTI induced performance degradation. This estimate may not be correct always due to unpredictable temperature, signal activity or process variations. So an active on-chip reliability sensor is introduced. An accurate detection of reliability degradation is possible and the detected signals can be used for corrective actions such as an optimal body bias signal to avoid timing failures. The additional design overhead introduced by body biasing and sensor circuits includes increased area, power and interconnections.

7.4. Stand-by V_{DD} scaling.

The primary design consideration in memory array is to minimize cell area and improve stability. It is a very common practice to reduce the

V_{DD} of a memory cell in standby mode to reduce power consumption. This V_{DD} scaling can be utilized to reduce the NBTI degradation because the NBTI is dependent on the vertical electric field at the Si-SiO_2 interface. When the activity factor of memory is lower, i.e. more time in standby mode, the V_{DD} scaling can significantly improve the stability of memory cell. It is advantageous that this technique can be easily incorporated to the existing memory structures without much additional overhead.

8. Conclusion

This paper acknowledges NBTI as a major reliability challenge and emphasizes the need for reliability aware circuit design by pointing out the severity of NBTI degradation in random logic circuits and SRAM memory cells. Apart from NBTI, its counterpart PBTI also gains importance in the promising Hf based high-k dielectrics. The transistor reliability will be a severe problem in future technology nodes which makes the device life time shorter than predicted and thus reliability aware CMOS integrated circuit design will move up the curve as a promising research field for the future CMOS technologies which ensures robust and stable products.

References

[1] Sang Phill Park, Kaushik Roy, Kunhyuk Kang, "Reliability Implications of Bias Temperature Instability in Digital ICs", IEEE Design & Test of Computers, Nov/Dec 2009, 0740-7475/09

[2] M.A. Alam, S. Mahapatra, "A comprehensive model of PMOS NBTI degradation", Science Direct, Microelectronics Reliability 45 (2005) 71–81

[3] Dieter K. Schroder, "Negative bias temperature instability: What do we understand?", Science Direct, Microelectronics Reliability 47 (2007) 841–852

[4] S.V. Kumar, C.H. Kim, and S.S. Sapatnekar, "NBTI-Aware Synthesis of Digital Circuits," Proc. 44th Design Automation Conf. (DAC 07), ACM Press, 2007, pp. 370-375