An analytical model for read static noise margin including soft oxide breakdown, negative and positive bias temperature instabilities

Behrouz Afzal a, Behzad Ebrahimi a, Ali Afzali-Kusha a,⇑, Hamid Mahmoodi b

a Nanoelectronics Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, College of Engineering, Tehran, Iran
b Department of Electrical and Computer Engineering, San Francisco State University, CA, USA

1. Introduction

Reliability concerns have slowed down the operating voltage reduction with each new technology node while the gate oxide has become thinner for a given dielectric material [1]. This results in larger electric fields in the gate oxide which could form traps in the oxide. As the number of traps increases, they start to overlap forming a conduction path between the gate and channel which may create a gate tunneling current. This type of breakdown which is known as soft oxide breakdown (SBD) may occur in both SiO2 and high-κ gate dielectrics [2,3]. The SBD phenomenon can deteriorate the functionality of SRAM cells [4–6].

Larger vertical electric fields can cause another reliability problem for transistors which is called bias temperature instability (BTI) [7]. For PMOS transistors, when negative gate biases are applied, high energy holes break Si–H bonds at the Si–SiO2 interface forming interfacial traps. These traps increase the threshold voltage of the device, and hence, affecting the performance of the transistor (NBTI effect) [7]. Similarly, in the case of NMOS transistors with high-κ, under positive gate biases, a significant charge trapping may occur increasing the threshold voltage with time (PBTI effect) [8].

The effect of NBTI and PBTI on the stability and performance of SRAM cells have been investigated in [4–5,10–16]. In particular, the work presented in [12] studies the effect on the read stability for a 90 nm technology. The study shows that SBD is a dominant factor in the read stability degradation. In [13], a read SNM model which considers SBD along with NBTI was proposed. The model was based on the simple square law model for the I–V characteristic and did not consider the PBTI effect.

The contributions of this paper are as follows:

- We use a more accurate expression for the I–V characteristic to present a more tangible read SNM model for smaller technologies.
- The proposed model considers the SBD effect as well as both PBTI (important for transistors with high-κ gate dielectrics) and NBTI effects.

The reminder of the paper is organized as follows. In Section 2, the models used for the SBD, NBTI, and PBTI effects are described while, in Section 3, a read SNM model considering the SBD effect with short channel I–V models is derived. In Section 4, the accuracy of the model is investigated by comparing its results to those of HSPICE. In addition, using this model, the read SNM degradation in the simultaneous presence of NBTI, PBTI, and SBD effects is studied. Finally, Section 5 concludes the paper.
Here, \( I_0 \) is the initial oxide tunnel current, \( t \) is the elapsed time, and \( GR \) is the defect current growth rate which has an exponential relation with the stress voltage and oxide thickness as

\[
GR = K_1 \exp(\theta_1 V_s - \theta_2 T_{ox})
\]

where \( V_s \) is the gate stress voltage, \( T_{ox} \) is the gate oxide thickness, and \( \theta_1, \theta_2, \) and \( K_1 \) are constants which may be found from experimental data [12].

The breakdown may occur between the gate and diffusion region (source or drain) or between the gate and the channel of the transistors. Experimental data shows that the pull down source breakdown causes more severe stability degradation in a conventional 6T SRAM cell which is shown in Fig. 1 [17], and hence, similar to [13], only the gate-source breakdown of the pull down transistor under stress (NR in Fig. 1) is considered in this work. Using Eq. (1), the SBD is modeled as a resistor \( (R_{SBD}) \) given by [12]:

\[
R_{SBD} = \frac{V_{ds}}{I_{ds}} \exp(-\beta GR)
\]

where \( V_{ds} \) is the supply voltage.

Next, we should model the NBTI and PBTI effects which increase the magnitudes of the threshold voltages of PMOS and NMOS transistors, respectively, as a function of time (t). The increase, which is denoted by \( \Delta V_{th} \), may be modeled by the DC reaction–diffusion (RD) framework as [7]

\[
|\Delta V_{th}| = K_{RD} t^\alpha
\]

Here, \( K_{RD} \) is a constant which depends on the gate–source bias \( (V_{gs}) \), temperature, and other technology parameters. Fig. 2 shows the change in \( V_{th} \) due to NBTI and PBTI using the reaction diffusion framework which has been calibrated with published data for a 32 nm technology node [18]. The results are for poly (oxide/polysilicon) and high-\( k \) (high-\( k \)/metal gate) cases. While the PBTI effect on the threshold voltage is dependent on both dielectric and gate materials, the threshold voltage change due to the NBTI effect is only dependent on the dielectric material [18]. The results reveal that the PBTI can be ignored for the case of poly gate while it should be considered for the case of high-\( k \).

### 3. Read SNM modeling considering SBD

Next, we should model the read SNM of the 6T SRAM cell during the read operation. For this purpose, we use the circuit which is shown in Fig. 1 where static noise sources \( (V_n) \) and soft oxide breakdown resistance \( (R_{SBD}) \) have been added. The resistance \( R_{SBD} \) which models the gate-to-source soft breakdown of NR is modeled as a linear resistor between the gate and the source. The resistance could degrade the read SNM considerably, and hence, we have considered only this scenario of breakdown in the analysis. The same approach has been taken in [13]. In addition, since the chance of multiple breakdown events is quite low, we ignore them in this work.

To improve the read SNM modeling approach presented in [13], we employ a more accurate expression for the \( I-V \) characteristics. Instead of the square law model, in this work, we use the BSIM3v3 short channel model equations for hand calculations [19]. In the linear region \( (V_{ds} < V_{dsat}) \), the drain current \( (I_{ds}) \) is given by [19]:

\[
I_{ds} = \beta \frac{1}{1 + \frac{V_{ds}}{V_{gs} - V_{th}}} \left( V_{gs} - V_{th} - \frac{A_{bulk} V_{ds}}{2} \right) V_{ds}
\]

where \( A_{bulk} \) is the bulk charge coefficient, \( L \) is the channel length, \( E_{sat} \) is the minimum electric field for the onset of velocity saturation, \( V_{gs} \), \( V_{ds} \), and \( V_{th} \) are the gate-source, drain-source, and threshold voltages, respectively, and

\[
\beta = \mu_{eff} C_{oxe} W
\]

where \( \mu_{eff} \) is the effective mobility, \( C_{oxe} \) is the gate capacitance per unit area, and \( W \) is the channel width. In the saturation region \( (V_{ds} > V_{dsat}) \), the current–voltage characteristics is expressed as [19]

\[
I_{ds} = \frac{\beta}{2 A_{bulk} + 1} \left( V_{gs} - V_{th} \right) \left( V_{gs} - V_{th} \right) \left( 1 + \frac{V_{ds} - V_{dsat}}{V_{A}} \right)
\]

where

\[
V_{dsat} = \frac{E_{sat} L \left( V_{gs} - V_{th} \right)}{A_{bulk} E_{sat} L + V_{gs} - V_{th}}
\]

Assuming PR is off, one may write the KCL equation at the node R as

\[
I_{NR} = I_{AR}
\]

Also, we suppose that NR and AR operate in the linear and saturation regions, respectively, and hence, we can write

\[
I_{NR} = \beta_{NR} \frac{1}{1 + \frac{V_{ds} - V_{th} - V_{th} - V_{th}}{E_{sat} L (E_{sat} - L)}} \left( V_{gs,nr} - V_{th} - \frac{A_{bulk} V_{ds}}{2} \right) V_{ds,nr}
\]

\[
I_{AR} = \frac{\beta_{AR}}{2 A_{bulk} + 1} \left( V_{dd} - V_{ds,nr} - V_{th,ar} \right) \left( V_{dd} - V_{ds,nr} - V_{th,ar} \right) \left( 1 + \frac{V_{dd} - V_{ds,nr} - V_{ds,nr} - V_{th,ar}}{V_{A}} \right)
\]
It has been shown that the transfer characteristics of $V_{gs-NR} - V_{ds-NR}$ have a fairly constant slope around its operating point where $NR$ is in the linear region [20]. The linear approximation of this characteristic may be expressed as [20]

$$V_{ds-NR} = V_0 - k V_{gs-NR}$$  \(12\)

The parameters in Eq. (12) may be found by fitting the model predictions to the simulation results [13]. However, these parameters are sensitive to the threshold voltage of $NR$ which is increased by PBTI. For the study of the PBTI effect, we need to find analytical models for these parameters as a function of the threshold voltage.

As the first step in finding the parameters $k$ and $V_0$, let us take partial derivatives of both sides of Eq. (9) with respect to $V_{gs-NR}$ and $V_{ds-NR}$.

$$\frac{\partial V_{gs-NR}}{\partial V_{gs-NR}} dV_{gs-NR} + \frac{\partial V_{ds-NR}}{\partial V_{gs-NR}} dV_{ds-NR} = \frac{\partial V_{gs-NR}}{\partial V_{ds-NR}} dV_{gs-NR} + \frac{\partial V_{ds-NR}}{\partial V_{ds-NR}} dV_{ds-NR}$$

Using Eq. (12), one may write

$$\frac{dV_{ds-NR}}{dV_{gs-NR}} = -k$$  \(15\)

Also, from Eq. (14), one may write

$$k = \frac{\left(\frac{\partial V_{gs-NR}}{\partial V_{ds-NR}} - \frac{\partial V_{gs-NR}}{\partial V_{gs-NR}} \right)}{\frac{\partial V_{ds-NR}}{\partial V_{ds-NR}}}$$  \(16\)

The above equation (Eq. (16)) is valid at all points in the operation region of our interest, and hence, may be solved at any arbitrary point in this region. Let us consider $V_{gs-NR} = V_{dd}$ where the corresponding $V_{ds-NR}$ is denoted by $V_{dd}$, and is small, which can be found from Eq. (9) by equating $V_{gs-NR}$ and $V_{ds-NR}$ equal to $V_{dd}$ and $V_{dd}$ respectively. Using proper approximations, one may write

$$(V_{dd} - V_{th-NR}) - \frac{1}{2} \frac{A_{bulk-N}}{A_{bulk-N}} V_{dd} = 0$$

which is a second order equation with respect to $V_{dd}$. Using $V_{gs-NR}$ and $V_{ds-NR}$, one can use Eqs. (10), (11), and (16) to find $k$. Finally, $V_0$ is found from Eq. (12).

Similarly, assuming $AL$ is off, one may write the KCL equation at the node L as

$$f = I_{NL} + V_{gs-NL}/R_{SBD} - I_{PL} = 0$$  \(18\)

Considering the fact that $NL$ and $PL$ operate in the saturation and linear regions, respectively, we have

$$I_{NL} = \frac{\beta_{NL}}{2A_{bulk-N}} \left(1 + \frac{V_0 + V_{ds-NR} - V_{th-NL}}{A_{bulk-N} E_{val} L_{NL}} \right) \left(\frac{1}{V_{dd} - V_{th-NR}} \right)^2$$

rewriting Eq. (27) using the parameters $a$ and $b$ yields

$$SNM = SNM(R_{SBD} = \infty) - \frac{a}{R_{SBD}} - \frac{b}{2 R_{SBD}^2}$$  \(28\)

where
The derivative of SNM with respect to $R_{SBD}$ can be easily found from Eq. (32). Note that in using Eq. (32), $V_n$ is equal to SNM, $V_{g_{s-NR}}$ is obtained from setting the delta of the quadratic equation (Eq. (23)) equal to zero, and $V_{g_{s-NR}}$ is obtained from Eq. (12). The first term in Eq. (28), which represents the SNM without considering the SBD effect, may be obtained using existing accurate models for the SNM which are functions of the threshold voltages of the transistors (see, e.g., [22–24]). For these models, to include the BTI effects and the SBD simultaneously, we may be used in our model (see, e.g., [12]). Thus, the accuracy of the proposed model is assessed by comparing the model predictions with those of HSPICE simulations for 22 and 32 nm technology [21]. First, we present the results for the read SNM versus $R_{SBD}$ for the nominal threshold voltage as well as 5% and 10% threshold voltage shifts of both the transistors PL and NR. The results which are shown in Fig. 4 reveal a very good accuracy for the model. Therefore, this hints that the model may provide a fast yet accurate estimation of the SNM compared to the simulation method.

In [13], it was shown that both the SBD and NBTI effect should be considered together to obtain the correct model for predicting the variation of the read SNM over the time. In Fig. 5, we have presented similar results with the difference of including PBTI too. The figure shows the difference of the read SNM changes for the case of considering both the BTI effects and the SBD simultaneously and the case where SBD and the NBTI/PBTI effects are considered separately. In the latter case, the change was obtained by adding the read SNM reduction due to the SBD effect when $\Delta V_{th} = 0$ and the NBTI/PBTI effects when $R_{SBD} = \infty$. Note that the results plotted in Fig. 5 are versus $R_{SBD}$ and obtained assuming that the increases in the MOS threshold voltages due to the BTI effects were 5% and 10%. As the figure indicates, when $R_{SBD}$ becomes smaller (stronger SBD effect) and threshold voltage change becomes larger (severe BTI effects), the difference becomes larger. NBTI/PBTI phenomena worsens the effect of SBD on the read SNM by 16% (8%) and 31% (34%) for the 5% (10%) increase in threshold voltages due to the BTI effects for 22 and 32 nm technologies, respectively. The same behavior was observed in [13].

This behavior is due to the fact that the change in the SNM value due to SBD is a function of the threshold voltage shifts [13]. Because the coefficients for the SBD terms ($a$ and $b$) are themselves functions of the threshold voltages, and hence, for a better accuracy both the NBTI and PBTI effects along with the SBD effect should be considered together. Note that the slope in Fig. 3 ($\lambda$ in Eq. (29)) becomes more negative when the threshold voltage change becomes larger, and consequently, $a(-\lambda)$ becomes higher. Thus, ASNM due to SBD becomes more which is also apparent from Eq. (28) (in which the second term becomes larger). It is also apparent from Eq. (28) that ASNM due to SBD is more sensitive to $a$ (MOS threshold voltages) as $R_{SBD}$ decreases. In Fig. 6, we have drawn the results of the model and HSPICE simulations for the variations of $\lambda$ as a function of the threshold voltage change of PL and NR for the two technologies. This graph demonstrates that increasing the BTI effects makes the SBD effect more detrimental on SNM.

Next, we plot the read SNM as a function of the stress time for the 22 and 32 nm technology. For this graph, we used the data in Fig. 2 for the $V_{nr}$ drift due to NBTI and PBTI for high-k, Eq. (3) for the SBD resistance calculation, and three arbitrary values of $GR$ as 1.6, 3.2, and 6.4 $\times 10^{-2}$/s (we did not have access to industrial data which is process dependent). As shown in Fig. 7, different growth rates result in different times for the onset of substantial change in the read SNM.

We also study the effect of the supply voltage on the read SNM degradation. For this purpose, the read SNM values for different supply voltages versus time have been shown in Fig. 8. For these results, the changes of $V_{th}$ due to NBTI and PBTI for the supply voltages of 0.7, 0.8 and 0.9 V have been obtained from the analytical expressions given in [25] and the SBD resistance from Eq. (3) by assuming that $GR$ increases 5 dec/V with $V_{th}$ [12]. The results show that while for smaller supply voltages the initial SNM value is lower, the rate of the SNM degradation due to NBTI/PBTI and SBD is smaller too. The accuracy degrades as the supply voltage scales more. This may be due to the assumption that all transistors operate in on state, but in lower supply voltages some transistors may...
approach operating in subthreshold region. Even for $V_{dd} = 0.7$ V, the error is very small (the mean error for $V_{dd}$ equal to 1, 0.9, 0.8 and 0.7 V is 1.1%, 1.5%, 1.9% and 4.5%, respectively).

Finally, Fig. 9 compares the Cumulative Distribution Function (CDF) of the read SNM versus minimum read SNM (RSNM0) obtained using 15,000 HSPICE Monte Carlo simulations and our proposed model under process variations and aging effects (NBTI/PBTI and SBD) for the 32 nm technology. We consider the threshold voltage of transistors due to process variations as Gaussian random variables [26]. The $3\sigma$ of the threshold voltages were set to 20% of their nominal values [22]. The CDF value at each RSNM0 shows the percentage of the cells whose read SNM values are smaller than RSNM0. The figure shows, for example, if the target read SNM is assumed to be 30 mV, the percentage of the cells with smaller read SNM values than 30 mV are about 0.15% and 22% initially and after 1 year, respectively. The comparison reveals a very good accuracy for the model which is evaluated in a very short period of time due to its analytical nature. Therefore, when we study the impact of process variations and aging on SRAM cell, using the proposed model is a very efficient method of calculating the read SNM compared to HSPICE simulations.
In this work, we proposed an analytical read SNM model which considered the soft oxide breakdown effect using an accurate model for the $I-V$ characteristic. Both the NBTI and PBTI effects were considered in the model by including the change of the threshold voltages in the SNM model. The accuracy of the model was verified by comparing its prediction with those of HSPICE simulations for 32 and 22 nm technologies. The comparison revealed a very good accuracy for the model in these technologies. The accuracy of the model was verified for a wide range of supply voltages. The model can be used for the prediction of the minimum supply voltage which should be used for having a target yield when the lifetime is given. In addition, the results showed that both NBTI and PBTI phenomena worsen the SBD effect on the read SNM. This suggests that the effect of NBTI/PBTI and SBD should be included in the model simultaneously.

**5. Conclusion**

In this work, we proposed an analytical read SNM model which considered the soft oxide breakdown effect using an accurate model for the $I-V$ characteristic. Both the NBTI and PBTI effects were considered in the model by including the change of the threshold voltages in the SNM model. The accuracy of the model was verified by comparing its prediction with those of HSPICE simulations for 32 and 22 nm technologies. The comparison revealed a very good accuracy for the model in these technologies. The accuracy of the model was verified for a wide range of supply voltages. The model can be used for the prediction of the minimum supply voltage which should be used for having a target yield when the lifetime is given. In addition, the results showed that both NBTI and PBTI phenomena worsen the SBD effect on the read SNM. This suggests that the effect of NBTI/PBTI and SBD should be included in the model simultaneously.