Negative Bias Temperature Instability in CMOS Devices


Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
*School of Electrical Engineering and Computer Science, Purdue University, West Lafayette, IN, USA
Tel: 91-22-25720408 email: souvik@ee.iitb.ac.in

Abstract

This paper reviews the experimental and modeling efforts to understand the mechanism of Negative Bias Temperature Instability (NBTI) in p-MOSFETs, which is becoming a serious reliability concern for analog and digital CMOS circuits. Conditions for interface and bulk trap generation and their dependence on stress voltage and oxide field, temperature and time are discussed. The role of inversion layer holes, hot-holes and hot-electrons are also discussed. The recovery of generated damage and its bias, temperature and AC frequency dependence are discussed. The degradation and recovery is modeled using the standard Reaction-Diffusion theory, and some unique data scaling features are pointed out. The impact of gate-oxide nitridation is also reviewed.

Keywords: MOSFET; NBTI; interface traps; bulk traps; parameter instability; Reaction-Diffusion model; hydrogen diffusion

1. Introduction

Negative Bias Temperature Instability (NBTI) of p-MOSFET parameters (threshold voltage, linear and saturation drain current, gate-drain capacitance, etc.) is becoming a serious reliability concern for digital and analog CMOS circuits (see [1] and references therein). After extensive characterization it has been found [1-7] that NBTI (i) shows a t^n power law time dependence with n~0.17-0.3, (ii) is a temperature activated process with an activation energy (E_A) of 0.1-0.2eV, (iii) is more dominant for p-MOSFETs under inversion, (iv) anneals if stress is removed, (v) is reduced if deuterium anneal is used for passivation in place of hydrogen, (vi) is influenced by the presence of water, boron, and nitrogen in the gate oxide, and (vii) is due to either (or both) interface (N_{IT}) and bulk-trap (N_{OT}) generation.

There have been some efforts to model NBTI as well [8-10]. However, the basic NBTI mechanism is not yet fully understood. Moreover, NBTI has not been carefully studied for p-MOSFETs under non-zero substrate bias (V_B). This is an important concern since many circuits utilize V_B to control the device threshold voltage (V_T), (e.g., for multiple V_T CMOS, standby leakage reduction, etc.). As of today, the issues of interest are: (i) generation of N_{IT} and N_{OT}
and their gate bias ($V_G$), $V_B$, temperature (T), and time (t) dependence; (ii) recovery of $N_{IT}$ and $N_{OT}$ and their $V_G$, T and AC frequency (f) dependence, and (iii) impact of nitrogen on $N_{IT}$ and $N_{OT}$. Proper understanding of the complete NBTI mechanism will help determine (i) reliability budget for a technology node, (ii) burn-in and test conditions, (iii) TCAD and SPICE models, and (iv) process parameters for NBTI control. This paper attempts to develop a consistent picture of the NBTI phenomenon under a wide range of operating conditions, and identifies possible areas where the present understanding is incomplete and requires further attention.

2. Proper choice of stress gate bias

Fig.1 shows the time evolution of $\Delta V_T$ and SILC for $T_{PHY}=36\,\AA$ devices under different stress $V_G$. For low stress $V_G$, $\Delta V_T$ shows the well-known power law in time with exponent 0.25. For high stress $V_G$, $\Delta V_T$ increases at longer time and shows much larger time exponent. SILC is not seen for low stress $V_G$ but is observed for high stress $V_G$. The onset (in time) and magnitude of SILC coincides with that of enhanced $\Delta V_T$ at high stress $V_G$. It is obvious that such long-time $\Delta V_T$ increase at high stress $V_G$ seriously affects extrapolated lifetime and related voltage acceleration analysis, and hence calls for careful attention.

Fig.2 shows p-MOSFET energy band diagram in inversion at low and high $V_G$. At low $V_G$, electrons tunnel from gate and holes tunnel from substrate. At high $V_G$, electrons tunnel from gate and cause impact ionization at the substrate and thereby generate hot holes (HH). These generated HH inject back into the oxide and is responsible for $N_{OT}$ generation [4,11].

Note that in general, $\Delta V_T$ is due to both $N_{IT}$ and $N_{OT}$. Since the appearance of SILC indicates $\Delta N_{OT}$ (as SILC monitors mid-oxide traps) [11], it is clear that long-time enhanced $\Delta V_T$ at high stress $V_G$ is due to (hole trapping in) $\Delta N_{OT}$ [4]. This is verified by the calculated $\Delta N_{OT}$ contributions (shown by the lines in Fig.1). As $\Delta N_{OT}$ is a strongly $V_G$ activated process, it becomes negligible compared to $\Delta N_{IT}$ at lower stress $V_G$. Therefore, NBTI must always be evaluated at low to moderate $V_G$ to avoid impact ionization, HH generation and undesirable contribution from $\Delta N_{OT}$.

As a proof, Fig.3 plots $\Delta V_T$ versus $\Delta N_{IT}$ for a wide range of stress $V_G$. The maximum $V_G$ was kept low to avoid HH generation in this device. Excellent correlation over a wide range of $V_G$ suggests the absence of $\Delta N_{OT}$. Otherwise, strong $V_G$ acceleration of $\Delta N_{OT}$ would result in higher $\Delta V_T$ for a given $\Delta N_{IT}$ with increasing $V_G$. Therefore, NBTI is only due to interface trap generation for lower (proper) stress $V_G$.

As a guideline for choosing maximum stress $V_G$, Fig.4 shows measured gate ($I_G$), substrate ($I_B$) and source/drain ($I_{SD}$) current as a function of $V_G$ using carrier separation measurement. According to Fig.2, $I_G=I_B+I_{SD}$ for low stress $V_G$, where $I_G$ and $I_{SD}$ are respectively due to electrons and holes tunneling.
from gate to substrate and source/drain to gate. For high stress $V_G$, $I_b = I_g + I_{SD}$ and the holes generated out of impact ionization flow out of source/drain and cause a sign reversal for $I_{SD}$. The “dip” in $I_{SD}(V_G)$ characteristics indicates significant presence of HH. $V_G$ for NBTI stress must be chosen below this point to ensure negligible $\Delta N_{IT}$ and $\Delta V_T \sim \Delta N_{IT}$ only.

3. Factors influencing interface-trap generation

Fig. 5 shows correlation of $\Delta V_T$ and $\Delta D_{IT}$ for a wide stress $V_G$ range on devices having thermal, non-nitrided oxide. Maximum $V_G$ kept low to avoid impact ionization. $\Delta D_{IT} (=\Delta N_{IT}/\Delta E)$ obtained from charge pumping, $\Delta E$ (energy zone of recombination) determined in pre-stress from multi-frequency measurements. Data from [5].

Fig. 4. Gate, substrate and source/drain current as a function of $V_G$ from carrier separation measurements under $V_B=0V$.

Fig. 3. Correlation of $\Delta V_T$ and $\Delta D_{IT}$ for a wide stress $V_G$ range on devices having thermal, non-nitrided oxide. Maximum $V_G$ kept low to avoid impact ionization. $\Delta D_{IT} (=\Delta N_{IT}/\Delta E)$ obtained from charge pumping, $\Delta E$ (energy zone of recombination) determined in pre-stress from multi-frequency measurements. Data from [5].

Fig. 6 also shows normalized $\Delta I_D$ for n-MOSFET inversion (NINV) stress under identical $E_{OX}$ as PINV stress. PINV shows much larger $\Delta N_{IT}$ than NINV, indicating possible role of holes. Note, availability of holes at the Si/SiO$_2$ interface is more for PINV stress (inversion layer) compared to NINV stress (tunneling from gate). The role of holes is further verified in Fig. 7, which shows $\Delta V_T$ as a function of inversion hole density for different $T_{PHY}$. Excellent correlation over a wide range of $T_{PHY}$ again indicates that holes play an important role behind $N_{IT}$ generation.

Fig. 6. Time evolution of normalized $I_D$ degradation for n- and p-MOSFETs (thermal, non-nitrided oxide) under different stress configurations. Data from [5].

Fig. 8 plots $\Delta V_T$ versus energy dissipated by hot electrons (tunneling from gate) at Si/SiO$_2$ interface for different $T_{PHY}$ devices. The absence of correlation clearly shows that hot electrons do not play a role behind $\Delta N_{IT}$. Note that the absence of $\Delta V_T$ versus $V_G$ correlation (Fig. 5) also suggests that hot electrons do not play a role behind $\Delta N_{IT}$ even when their energy is
considered in the ballistic limit \((E \sim q \cdot V_G)\). All the above results clearly show that \(\Delta N_{IT}\) during NBTI stress is governed by \(E_{OX}\) and inversion holes.

\[
\begin{align*}
\text{Si-H} + h & = \text{Si}^+ + H^0, \quad H^0 + H^0 = H_2 \\
\text{Si-H} + h + H^0 & = \text{Si}^+ + H_2
\end{align*}
\]

(1)

Time evolution of \(\Delta N_{IT}\) is governed by the following coupled differential equations [8-10]:

\[
\begin{align*}
dN_{IT}/dt & = k_F [N_0 - N_{IT}] - k_R N_{H}(x=0) N_{IT} \\
dN_{H}/dt & = D d^2N_{H}/dx^2
\end{align*}
\]

(2)

where \(k_F, k_R\) are the forward and reverse reaction rates, \(N_0\) is the total density of Si-H bonds, \(N_H\) and \(D\) respectively are the concentration and diffusivity of released H species. The R-D model and its solution are explained in Fig. 9.

4. Model for interface-trap generation

\(N_{IT}\) generation during NBTI has been explained using the Reaction-Diffusion (R-D) model [8]. The model states that \(\Delta N_{IT}\) is due to the dissociation of Si-H bonds at the Si/SiO\(_2\) interface (reaction) and the subsequent movement of released H species away from the interface (diffusion) which leaves behind Si- (\(N_{IT}\)). The model in its basic form does not say how the reaction takes place, or for that matter what is the nature (neutral/charged) of the diffusing species.

Though the exact reaction process is unknown, it is speculated that inversion layer holes tunnel into the oxide and are captured by Si-H bonds. This makes the Si-H bonds weak, which are subsequently broken by thermal excitation or otherwise. Several reaction pathways are possible (\(h = \) holes, \(H^+ = \) ionic, \(H^0 = \) atomic and \(H_2 = \) molecular hydrogen) [10]:

\[
\begin{align*}
\text{Si-H} + h & = \text{Si}^+ + H^0, \quad H^0 + H^0 = H_2 \\
\text{Si-H} + h + H^0 & = \text{Si}^+ + H_2
\end{align*}
\]
condition phase-3 (and beyond) are experimentally observed. Recently, the coupled R-D equations have been numerically solved [9,10]. It has been shown that the power law time exponent in phase-3 depends on the nature of the diffusing species [10], as listed in Table-I. Typical H diffusion profile as obtained using numerical simulation (for H0) is shown in Fig. 10.

### Table-I. Simulated power law time exponent in phase-3 for different diffusing species [10].

<table>
<thead>
<tr>
<th>Species</th>
<th>Exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0</td>
<td>0.25</td>
</tr>
<tr>
<td>H2</td>
<td>0.165</td>
</tr>
<tr>
<td>H+, H2</td>
<td>0.165 – 0.25</td>
</tr>
<tr>
<td>H+, H2, H+</td>
<td>0.165 – 0.5</td>
</tr>
<tr>
<td>Trap or Release</td>
<td>Increase or decrease</td>
</tr>
</tbody>
</table>

5. Universal scaling scheme (R-D model)

In phase 3, a simple analytic solution of the R-D equations yields [4,9]:

$$\Delta N_{IT} = \left[ \frac{k_F N_0}{k_R} \right]^{1/2} (Dt)^n = S_N (Dt)^n \quad (3)$$

By assuming identical $E_A$ for $k_F$ and $k_R$ and a neutral H species, $S_N$ and D become only dependent of $E_{OX}$ and D respectively. This implies that $\Delta V_T$ (t) data measured at different $E_{OX}$ (fixed T) and different T (fixed $E_{OX}$) can be respectively scaled by $S_N$ (reaction term; only $E_{OX}$ activated) and D (diffusion term; only T activated), as illustrated by Fig. 11. Also note that if the assumption of $E_A (k_F) \sim E_A (k_R)$ is true, then (3) predicts that $E_A (NBTI) \sim E_A (D) \times n$ [7].

![Fig. 11. Scaling scheme for $\Delta N_{IT}$ versus t data. Density axis reflects higher Si-H bond dissociation. Time axis denotes faster hydrogen diffusion. According to [4].](image)

6. Experimental validation of R-D model

Fig. 12 shows $\Delta N_{IT}$ for stress at various $E_{OX}$ and T (low $V_G$, high T). $\Delta N_{IT}$ shows $t^n$ ($n \sim 1/5$) dependence at short time that increases at longer time. Break time ($t_{BREAK}$) decreases and the post-break slope ($S_{BREAK}$) increases with increase in T. However, $t_{BREAK}$ and $S_{BREAK}$ are insensitive to $E_{OX}$. This long-time $\Delta N_{IT}$ increase is unlike that observed at high $V_G$ (section 2), and is a diffusion related process. Based on R-D model prediction, long-time $\Delta N_{IT}$ enhancement is due to absorption of H species into poly-Si (absorbing poly-Si) once the H diffusion front hits the SiO2-poly-Si interface (at $t=t_{BREAK}$) [5]. As per Table-I, the time exponent n for $t < t_{BREAK}$ (phase-3, diffusion in oxide) suggests the H species as neutral H0 or H2.

Moreover as per the universal scaling scheme (section 5), $E_{OX}$ and T dependent data for $t < t_{BREAK}$ is scaled (see Fig. 12) along Y- and X-axis directions respectively to unique relations [4,7]. Y-axis scaling of varying $E_{OX}$ data yields field activation of (total) NBTI, X-axis scaling of varying T data yields $E_A$ of D, while Y-axis scaling of T dependent data (not shown) yields $E_A$ of (total) NBTI.

Fig. 13 (top) plots the T activation of $\Delta N_{IT}$, $1/t_{BREAK}$ and D. Obtained $E_A$ (D) suggests H species as neutral H2 [12]. $E_A$ for NBTI ($\Delta N_{IT}$) is related to $E_A$ (D) by the time exponent n ($\sim 1/5$), according to that predicted by the R-D model [7]. Furthermore, similar $E_A$ for D and $1/t_{BREAK}$ confirms that the T activated enhanced $\Delta N_{IT}$ is a diffusion related phenomenon. As a proof, Fig. 13 (bottom) plots $t_{BREAK}$ as a function of $T_{PHY}$. $t_{BREAK}$ reduces with
7. Recovery of interface traps

Fig. 14 shows the generation and recovery of $\Delta N_{IT}$ during and after NBTI stress. According to the R-D model, the forward reaction stops but the reverse reaction goes on once the stress is stopped. The H near the Si/SiO$_2$ interface rapidly anneals the broken Si- bonds, and creates a “diffusion hole”. This drives further H diffusion (from the oxide bulk) towards the Si/SiO$_2$ interface and further recovery of $\Delta N_{IT}$ takes place [9]. Fig. 14 also shows that the R-D model can efficiently predict $\Delta N_{IT}$ buildup and recovery during both stress and post-stress phases. The H diffusion profile (for H$^0$) during recovery is shown in Fig. 15.

Fig. 13. (top) Temperature activation of $1/t_{BREAK}$, $D$ (of H species in SiO$_2$) and $\Delta N_{IT}$ (NBTI activation). (bottom) $t_{BREAK}$ as a function of $T_{PHY}$. Data from [7].

Identical $\Delta N_{IT}$ recovery is observed for successive stress-recovery cycles, while recovery is very weakly dependent on post-stress $E_{OX}$ [6]. Of interest is $\Delta N_{IT}$ recovery measured at high T, especially after long-time stress in samples having absorbing poly-Si. Fig. 16 shows the magnitude of $\Delta N_{IT}$ generation and recovery and fractional recovery for 2000s stress and post-stress phases. The magnitude of $\Delta N_{IT}$ generation and recovery increases at larger T. However, the fractional recovery reduces at larger T (more $\Delta N_{IT}$ generation than recovery). This is due to H$_2$ loss or “lock-in” in poly-Si at higher T (due to faster D in poly-Si) [6], which results in less H$_2$ availability in the oxide for annealing the Si-bonds.

8. AC stress

Fig. 17 shows the pulse frequency dependence of $\Delta N_{IT}$. Degradation under AC is always lower than
that under DC NBTI [7,9,10]. \( \Delta N_{IT} \) is frequency independent for lower \( f \) and the reduction (w.r.t DC) is due to recovery effects. At higher \( f \), \( \Delta N_{IT} \) reduces further due to non-equilibrium reaction effects, which can also be predicted from R-D simulations [9,10]. Significant lifetime improvement can be achieved for circuits under switching conditions.

9. Effect of substrate bias

Fig. 18 (top) shows time evolution of \( \Delta V_T \) and \( \Delta N_{IT} \) for \( V_B=0V \) and \( V_B>0V \) stress. (bottom) Enhanced \( \Delta V_T, \Delta N_{IT} \), \( \Delta N_{OT} \) and SILC as a function of stress time. Thermal, non-nitrided oxide. \( \Delta X \) is \( \Delta N_{IT}, \Delta V_T \) or \( \Delta N_{OT} \) (as explicitly mentioned in the legend).

Fig. 19 shows \( \Delta N_{IT} \) generation (with and without \( V_B \)) and recovery for different post-stress bias. \( \Delta N_{IT} \) recovery is weakly dependent on post-stress bias. It is important to note that the amount of \( \Delta N_{IT} \) recovery is identical for stress with and without \( V_B \), i.e., \( V_B>0V \) stress induced enhanced \( \Delta N_{IT} \) does not recover.

Note, \( V_B>0V \) stress also causes impact ionization and HH generation [4]. HH creates \( \Delta N_{OT} \), which is due to broken Si-O bonds [11]. Mid-oxide broken Si-O bonds show up as SILC, while those near interface show up as additional \( \Delta N_{IT} \). Hence, additional \( \Delta N_{IT} \) shows identical power-law exponent as SILC (see Fig. 18), and does not recover (see Fig. 19). Higher degradation due to \( V_B>0V \) stress is a serious concern and must be carefully addressed.

10. Effect of nitrogen

Fig. 20 shows \( \Delta V_T \) for different amount of N\(_2\) content in the gate oxide. Data obtained from [13] for films grown by RTO and followed by RTN. \( \Delta V_T \) always increases with increasing N\(_2\) content (under identical stress condition), as also reported by others [14]. Though possible mechanism is yet unclear, it
has been suggested that $N_2$ reduces reaction energy [13] or creates hole-trapping centers [14], all of which can increase $\Delta V_T$. $N_2$ induced enhanced NBTI is the primary concern for future technology nodes and needs urgent attention.

11. Conclusions

The mechanism of NBTI is studied under a wide range of stress conditions. It is shown that high stress $V_G$ must be avoided to minimize unwanted additional contribution from HH induced $\Delta N_{OT}$. For “correct” stress, NBTI is purely due to $\Delta N_{IT}$ (in thermal, non-nitrided films) and is governed by inversion holes and $E_{OX}$ (not by $V_G$ and hot electrons). Generation and recovery of $\Delta N_{IT}$ is explained well with the R-D model, which shows that diffusion of H species plays an important role on time-rate of $\Delta N_{IT}$. A universal data scaling methodology is presented, the diffusion species is identified as neutral $H_0$ or $H_2$, and relation between $E_A$ of diffusion and $E_A$ of overall NBTI is shown. It is shown that faster $H_0$ or $H_2$ diffusion in poly-Si increases the long-time degradation and reduces post-stress recovery. It is shown that $\Delta V_T$ is larger for $V_B>0\text{V}$ stress because of additional contribution from $\Delta N_{OT}$. Finally, NBTI is shown to increase for nitrided gate oxide, though the exact mechanism still unclear.

References