Recent Issues in Negative-Bias Temperature Instability: Initial Degradation, Field Dependence of Interface Trap Generation, Hole Trapping Effects, and Relaxation

Ahmad Ehteshamul Islam, *Student Member, IEEE*, Haldun Kufluoglu, *Student Member, IEEE*, Dhanoop Varghese, *Student Member, IEEE*, Souvik Mahapatra, *Member, IEEE*, and Muhammad Ashraful Alam, *Fellow, IEEE*

(Invited Paper)

***Abstract*—Recent advances in experimental techniques (on-the-fly and ultrafast techniques) allow measurement of threshold voltage degradation due to negative-bias temperature instability (NBTI) over many decades in timescale. Such measurements over wider temperature range (−25 °C to 145 °C), film thicknesses (1.2–2.2 nm of effective oxide thickness), and processing conditions (variation of nitrogen within gate dielectric) provide an excellent framework for a theoretical analysis of NBTI degradation. In this paper, we analyze these experiments to refine the existing theory of NBTI to 1) explore the mechanics of time transients of NBTI over many orders of magnitude in time; 2) establish field dependence of interface trap generation to resolve questions regarding the appropriateness of power law versus exponential projection of lifetimes; 3) ascertain the relative contributions to NBTI from interface traps versus hole trapping as a function of processing conditions; and 4) briefly discuss relaxation dynamics for fast-transient NBTI recovery that involves interface traps and trapped holes.**

***Index Terms*—Fast transient recovery, field acceleration, hole–trapping, interface traps, negative-bias temperature instability (NBTI), reaction–diffusion (R-D) model, time exponent.**

I. Introduction

NEGATIVE-BIAS temperature instability (NBTI), as the name suggests, indicates a temperature-accelerated degradation in PMOS devices when it is stressed with negative gate voltage. Although NBTI was identified as a reliability concern in integrated circuits since the mid-1960s [1], its significance became particularly important below the 130-nm technology node [2]–[7], as gate oxide thickness was scaled below 2 nm. This transition is a reflection of two prevailing trends: 1) increasing oxide electric field by reducing the oxide thickness, without a corresponding reduction in supply voltage, for better transistor performance [8], and 2) use of oxinitride gate dielectric to reduce gate leakage and boron penetration effects [9]–[11]. NBTI results in variation of transistor parameters (e.g., threshold voltage, transconductance, saturation current, etc.) and is generally monitored through threshold voltage shift (ΔV_T), where

$$\begin{aligned}\Delta V_T &= \Delta V_{\rm IT} + \Delta V_H \\ &= \alpha \frac{q\Delta N_{\rm IT}(t)}{C_{\rm ox}} + \frac{\int_0^{T_{\rm ox}} \int_E x\rho_H(x,E,t)dE\,dx}{C_{\rm ox}T_{\rm ox}}. \end{aligned} \quad (1)$$

Here, $\Delta V_{\rm IT}$ and ΔV_H refer to the contributions to ΔV_T from interface traps ($N_{\rm IT}$) and hole trapping (ρ_H), respectively, $C_{\rm ox}$ is the oxide capacitance, $T_{\rm ox}$ is the oxide thickness, α accounts for the fraction of donor type [12] $N_{\rm IT}$ above the Fermi level, and $\rho_H(x,E,t)$ represents the trapped holes at location x (measured into the oxide from the poly/oxide interface) and at energy E (below the dielectric conduction band).

Based on the data collected over the years, it has been reported that NBTI shift in devices with pure SiO_2, plasma SiON (e.g., having $< 25\%$ atomic N_2 dose, determined by XPS [13] for EOT ~ 1.3 nm), or thin thermal SiON gate dielectrics (henceforth referred as type I devices) may be attributed to $N_{\rm IT}$ generation [first term in (1)] at the Si/SiO_2 interface. For these type I films, the followings have been suggested.

1) NBTI degradation mainly results from depassivation of Si–H bonds at the Si/dielectric interface [i.e., $\Delta V_T \cong \Delta V_{\rm IT}$ in (1)] and resultant diffusion of hydrogen species into gate dielectric and poly-Si [see Fig. 1(b)]. As such, several groups have used different versions of reaction–diffusion (R-D) model to interpret NBTI degradation [2], [4], [6], [7], [12], [14]–[23].

Manuscript received January 10, 2007; revised April 26, 2007. This work was supported by Applied Materials, TSMC, Renesas Technologies, and Semiconductor Research Corporation. The review of this paper was arranged by Editor D. Esseni.

A. E. Islam, H. Kufluoglu, D. Varghese, and M. A. Alam are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285 USA (e-mail: aeislam@purdue.edu).

S. Mahapatra is with the Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai 400076, India.

Digital Object Identifier 10.1109/TED.2007.902883

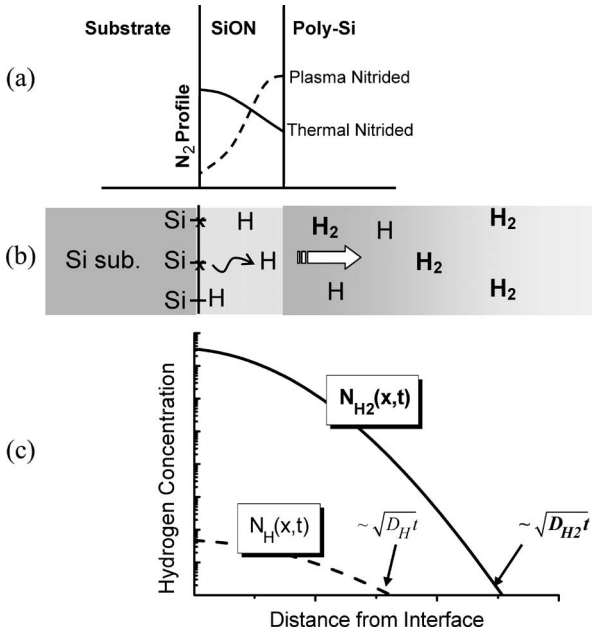


Fig. 1. (a) Typical nitrogen profile in plasma and thermal nitrided gate dielectric [13], [54]. (b) Schematic representation of Si–H bond dissociation and resulting diffusion of hydrogen species (H and H_2) through dielectric and poly-Si [17] during NBTI stress. (c) At long stress time, the diffusion of H_2 (concentration N_{H2}) dominates, and the concentration of H (N_H) is small.

2) At long stress time ($t_{\rm Stress} > 10-100$ s), ΔV_T shows power-law behavior when plotted with respect to time ($\Delta V_T \sim t^n$) with a consistent time exponent n of $\sim$1/6 (constant over several decades in time [19], [20]). The R-D model attributes this robust long-term $n \sim 1/6$ exponent to the diffusion of molecular hydrogen (H_2) [4], [7], [12]. However, the implication and experimental validation of the R-D model at a very short stress time is not studied.
3) This "H_2 R-D" model also provides a phenomenological interpretation of temperature and field dependencies of NBTI for long-term stress [7], [12], [15], [16]. However, the physical basis of the temperature and field dependencies has not been extensively explored.
4) The R-D model also anticipates essentially frequency-*independent* AC NBTI degradation (shown in [24], [25]), which has been observed in some [26], [27] but not all [28] experiments. The origin of this discrepancy is not fully understood.
5) NBTI recovery contaminates measured ΔV_T and should be avoided by using on-the-fly measurement (OTFM) [16], [29], [30] or ultrafast measurement (UFM) [19]. However, there is a perception in recent reports [6], [19] that fast transient recovery of NBTI is inconsistent with the classical R-D model [19], [31].

Again, for type II devices (e.g., having gate dielectrics like thick thermal SiON, high-K, etc.), it has been suggested that in addition to $N_{\rm IT}$ generation:

6) Hole trapping in preexisting defects plays an important role in determining overall time kinetics of NBTI degradation [28], [30], [32]–[35]. So far, these effects have been modeled by adapting the theories of thick oxides that consider trapping into the preexisting bulk traps but no detrapping from there [30]. [32], [35]. It is not known whether more detailed models for thin oxides, which account for both trapping and detrapping, would corroborate these conclusions.

Recently, 1) advances in measurement techniques (UFM [19], delay ~ 1 μs), 2) availability of NBTI data over broader temperature range (-25 °C to $+145$ °C) and for films of various N_2 concentrations and thicknesses [18], and 3) refinement of the R-D models [36], [37] provide us with an opportunity to explore both the *short-term* NBTI degradation and fast transient relaxation, as well as detailed physical mechanisms of the phenomenological parameters of both the R-D model and hole trapping kinetics in much greater detail than previously possible. We will systematically discuss these generalizations in this paper. Our discussion is divided into four parts.

1) *Time exponents*. For type I devices, the classical "H_2 R-D" model naturally interprets the robust long-term $n \sim 1/6$ power-law time exponent of NBTI. However, it cannot quantitatively explain the short-term ($<$ 1–10 s) degradation, particularly that obtained by UFM, with power-law dependence with a time exponent of $n > 1/6$ [19], [29], [32]. Quantitative explanation must consider gradual transformation [see Fig. 1(b)] of atomic to molecular hydrogen (H to H_2) rather than instantaneous transformation, as assumed in the "H_2 R-D" model [4], [12]. The resultant generalized R-D model is presented in both numerical and analytical formats in Sections II-A and B, respectively. In addition, different experimental groups have observed $n > 1/6$ region at a different OTFM timeframe [16], [19], [29], [32]. We show in Section II-C that this discrepancy results from the previously unappreciated role of *time-zero* delay for OTFM [37]. Finally, in addition to very short term transients in $<$ 10 s, we explore the possible origin of $n < 1/6$ (very soft saturation) at a very long time stress time in Section II-D.
2) *Field dependence*. The physical basis of the phenomenological field acceleration parameter γ in the R-D model, which was first demonstrated in [2], has never been fully explored. In Section III, we explore the origin of field dependence in the R-D model and show how it explains field-dependent NBTI data for type I devices, in addition to other experimental observations [18].
3) *Hole trapping*. We also consider the possible effect of hole trapping in NBTI degradation for type II devices, as considered in [28] and [32]–[34]. In these films, hole trapping has unique signatures on time exponents and activation energies [33]. We provide a consistent interpretation of these effects in Section IV.
4) *Recovery*. In Section V, we briefly comment on the origin of *logarithmic* time dependence in measured ΔV_T recovery for type I devices, which is recently reported in [6], [19], [28], within the context of the R-D model. This is a work in progress, and details of the theory will be published elsewhere; however, the issue is important enough to merit at least a brief discussion in this article.

II. Generalized R-D Model (for Type I Devices)

A. Numerical Analysis of NBTI Degradation, Including Short-Term Transient

The classical R-D model [7], [12] assumes that NBTI results from dissociation of Si–H bonds into positively charged Si dangling bonds and atomic H. These atomic H instantaneously dimerize (compared to relevant timescale of available long-term experimental data) into molecular hydrogen (H_2) before diffusing away from the interface. As one may suspect, such an assumption of instantaneous dimerization would not be appropriate to interpret very short term NBTI data. Such would require explicit consideration of H within the R-D framework, as this should be the first byproduct after interface trap generation, before getting transformed to H_2 [4], [12]. Although neutral charge state may not be a stable form of atomic hydrogen, its transient formation is indeed possible [38]. Therefore, diffusion of both H and H_2 as well as $H \leftrightarrow H_2$ conversion are explicitly incorporated in the generalized R-D framework by the following [36], [37]:

$$\frac{dN_{\rm IT}}{dt} = k_f(N_0 - N_{\rm IT}) - k_r N_{\rm IT} N_H^{(0)} \tag{2}$$

$$\frac{\delta}{2}\frac{dN_H^{(0)}}{dt} = D_H \frac{dN_H^{(0)}}{dx} + \frac{dN_{\rm IT}}{dt} - \delta k_H \left[N_H^{(0)}\right]^2 + \delta k_{H2} N_{H2}^{(0)} \tag{3a}$$

$$\frac{\delta}{2}\frac{dN_{H2}^{(0)}}{dt} = D_{H2} \frac{dN_{H2}^{(0)}}{dx} + \frac{\delta}{2} k_H \left[N_H^{(0)}\right]^2 - \frac{\delta}{2} k_{H2} N_{H2}^{(0)} \tag{3b}$$

$$\frac{dN_H}{dt} = D_H \frac{d^2 N_H}{dx^2} - k_H N_H^2 + k_{H2} N_{H2} \tag{3c}$$

$$\frac{dN_{H2}}{dt} = D_{H2} \frac{d^2 N_{H2}}{dx^2} + \frac{1}{2} k_H N_H^2 - \frac{1}{2} k_{H2} N_{H2}. \tag{3d}$$

Equation (2) represents the passivation/depassivation effects of Si–H bond, where k_f, k_r, N_0, $N_{\rm IT}$, and $N_H^{(0)}$ are defined as the Si–H bond-breaking rate, the Si–H bond-annealing rate, the initial bond density available before stress, the interface trap density, and the hydrogen density at the Si/dielectric interface, respectively. Equations (3a) and (3b) correspond to the conservation of fluxes of diffusing hydrogen species (H and H_2) near the interface (along x-axis), whereas (3c) and (3d) describe the diffusion (along x-axis) of H and H_2. $k_H N_H^2$ and $k_{H2} N_{H2}$ terms in (3) incorporate the $H-H_2$ conversion within the generalized R-D framework. Among the symbols used in (3), k_H and k_{H2} represent the generation and dissociation rates of H_2, D_H and D_{H2} represent the diffusion coefficients for H and H_2, N_H and N_{H2} represent the concentration of atomic and molecular hydrogen, and δ is the interfacial thickness ($\sim 1-2$ Å).

The solid lines in Fig. 2(a) show that the numerical solution of (2) and (3) interprets the data from [19] very well.[1] The solution reflects the dominance of $H \leftrightarrow H_2$ conversion during short-term stress (< 10 s) and extends the transition between the reaction-limited region ($n \sim 1$ [2], [12], [23]) and the H_2 diffusion-limited region ($n \sim 1/6$ [7], [12]) by several orders of timescale, which enables intuitive interpretation of short-term NBTI degradation. Fig. 3 shows the time-dependent formation of H and H_2 at the Si/SiO_2 interface and explicitly supports the notion that $H-H_2$ transformation governs NBTI generation at short timescales. Later, when $H \leftrightarrow H_2$ conversion reaches quasi-equilibrium (see Fig. 3), the dominance of H_2 diffusion results in $n \sim 1/6$. We use field dependence of k_f (see Section III) to interpret the voltage-dependent data in Fig. 2.

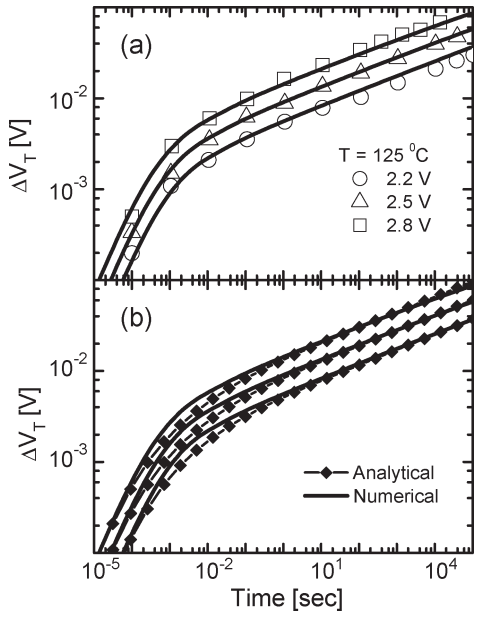


Fig. 2. (a) Simulation of the generalized R-D model ($H-H_2$ system) explains the experimental trends [19] at any stress time. R-D model parameters (consistent with literature [12], [17]): $k_f = 6 \times 10^{-3} E_c \exp(0.65 E_{\rm ox})$ s^{-1} (E_c: oxide electric field due to mobile carriers; $E_{\rm ox}$: total oxide electric field), $N_0 = 5 \times 10^{12}$ cm^{-2}, $k_r = 3 \times 10^{-9}$ $cm^3 s^{-1}$, $D_H = 3 \times 10^{-13}$ $cm^2 s^{-1}$, $D_{H2} = 1.8 \times 10^{-14}$ $cm^2 s^{-1}$, $k_H = 1.4 \times 10^{-3}$ $cm^3 s^{-1}$, and $k_{H2} = 95.4$ s^{-1}. (b) Comparison of (lines) numerical solution of (2) and (3) with (lines with symbols) the analytical solution of (7).

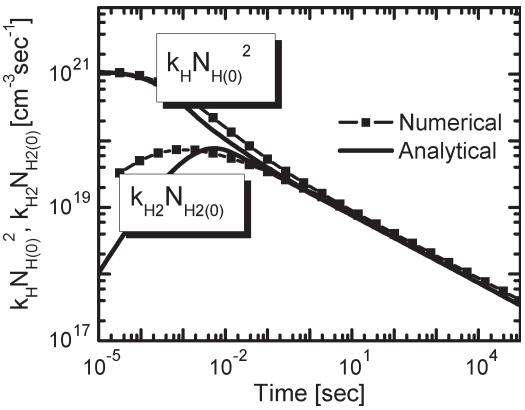


Fig. 3. Comparison of (lines) numerical results from (2) and (3) with (lines with symbols) analytical results for $k_H N_{H(0)}^2$ and $k_{H2} N_{H2(0)}$. Although analytical $N_{H2(0)}$ prediction is incorrect at short stress time (whereas $N_{H(0)}$ prediction is excellent), dominance of $k_H N_{H(0)}^2$ over $k_{H2} N_{H2(0)}$ still makes the analytical solution of $N_{\rm IT}$ appropriate for all stress time.

[1]Note that any uncertainty in measurement of voltage and/or time at very early phases may modify the fitting parameters used in Fig. 2. However, based on our unpublished results, we believe that the features are generally reproducible, and the parameters should be reasonably consistent across various experiments.

In discussing Fig. 2(a), it is important to note that the inability of the classical H_2 R-D model to reproduce the initial transient has been perceived as a fundamental limitation of the R-D model itself and has led some to speculate the existence of hole trapping at the initial stages of degradation [19]. Our analysis clearly shows that the limitation of the classical model can be traced to the (convenient and appropriate) assumption of instantaneous dimerization of H into H_2 to analyze long-term stress data but not to any inherent limitation of the R-D model itself.

B. Analytical Solution for NBTI Degradation

Although numerical models (2) and (3) adequately interpret the experimental data, an analytical model may be easier to use and provides additional insight. Assuming $N_0 \gg N_{\rm IT}$ [2], [12] and $dN_{\rm IT}/dt \sim N_{\rm IT}/t$ [23], (2) simplifies to

$$N_H^{(0)} = \frac{k_f N_0 - N_{\rm IT}/t}{k_r N_{\rm IT}}. \tag{4}$$

Moreover, the numerical solutions indicate that for continuous NBTI stress, $dN_H^{(0)}/dt$ and diffusion of H are negligible at all stress time so that (3a) reduces to

$$\frac{N_{\rm IT}}{t} = \delta k_H N_H^{(0)2} - \delta k_{H2} N_{H2}^{(0)}. \tag{5}$$

As stated earlier, H$-$H$_2$ diffusion reaches steady state at long $t_{\rm Stress}$, when H_2 diffusion dominates. Under that condition, simulation shows that

$$N_{\rm IT} \approx N_{H2}^{(0)} \sqrt{6 D_{H2} t}. \tag{6}$$

Eliminating $N_H^{(0)}$ and $N_{H2}^{(0)}$ from (4)–(6), we have

$$\frac{N_{\rm IT}}{t} - \frac{\delta k_H \left(k_f N_0 - N_{\rm IT}/t\right)^2}{k_r^2 N_{\rm IT}^2} + \frac{\delta k_{H2} N_{\rm IT}}{\sqrt{6 D_{H2} t}} = 0. \tag{7}$$

Equation (7) is the analytical expression of the H$-$H$_2$ R-D model presented in (2) and (3), and their comparison is presented in Figs. 2(b) and 3.[2] Fig. 2(b) shows that (7) 1) reduces to the reaction-limited solution ($N_{\rm IT} = k_f N_0 t$) when the first and third terms become negligible (compared to the second term) at a very short stress time; 2) reduces to the H_2 diffusion-limited solution $[N_{\rm IT} = (k_H/k_{H2})^{1/3} (k_f N_0/k_r)^{2/3} (6 D_{H2} t)^{1/6}]$ when $N_{\rm IT}/t$ becomes negligible at a long stress time; and 3) at an intermediate stress time, when the diffusive term (third term) in the equation (7) is negligible and $k_f N_0 \gg N_{\rm IT}/t$, a solution of $N_{\rm IT} = (k_f N_0/k_r)^{2/3} (\delta k_H t)^{1/3}$ provides a signature of dominant H$-$H$_2$ conversion [36]. Indeed, Fig. 2(b) suggests that the analytical solution of (7) is good enough to reproduce the numerical solution of (2) and (3) over $\sim$10 decades in stress time (maximum error of 20% in the transition region)! Moreover, a very good agreement between the numerical solution (solid lines) and the analytic solutions (dashed lines) in Fig. 3 suggests that (4) and (5) capture the dynamics of $N_H(x=0,t)$ and $N_{H2}(x=0,t)$ very well. The inaccuracy of N_{H2} at very early times is inconsequential because $N_{\rm IT}$ generation in this timescale is dictated by the generation and the transformation of N_H.

[2]A simple MATLAB code for evaluating (7) is given below: Nitcal = inline(['(Nit^3/t − delta*kh/kr^2*(kf*N0 − Nit/t)^2 − delta*kh2*Nit^3/sqrt(6*DH2*t))'], 'Nit', 't', 'kf', 'kr', 'kh', 'kh2', 'N0', 'DH2', 'delta'); Nit = fzero(Nitcal, [1 3*N0], options, time, kf, kr, kh, $kh2$, N0, DH2, delta).

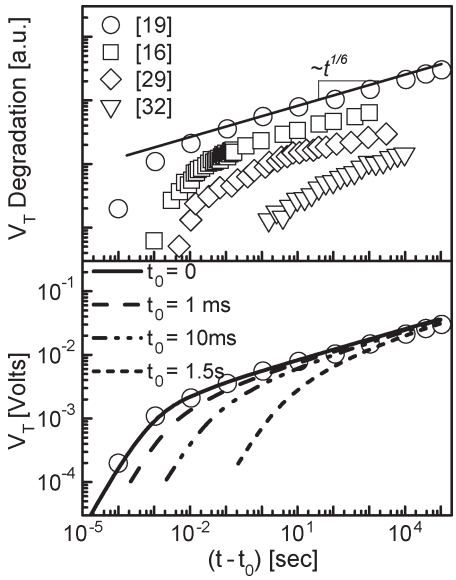


Fig. 4. (a) Experimental results from several groups [16], [19], [29], [32] show a widespread of stress time, at which the system reaches steady state with H_2 diffusion ($n \sim 1/6$). Here, datasets are arbitrarily shifted along the y-axis for clarity, i.e., relative magnitude between different datasets has no real significance. (b) Effect of *time-zero* delay in OTFM measurements can explain such spread in stress time, which indicates steady state. Here, $\Delta V_T(t - t_0) \sim |(N_{\rm IT}(t) - N_{\rm IT}(t_0))/N_{\rm IT}(t_0)|$. Simulation with $t_0 = 0$ is matched with experimental data from [19] at $V_{G,\rm Stress} = -2.2$V, 125 °C. As t_0 is increased, there is an increase in time exponent at a particular $(t - t_0)$ mainly in the short-term stress and also a rightward shift of the apparent transition point where the system reaches steady state with H_2 diffusion.

C. "Time-Zero" Delay in OTFMs

In the previous section, we have discussed how the H$-$H$_2$ R-D model can capture the experimental trends that are obtained using the UFM technique [19]. Fig. 4(a) shows short-term experimental data from a number of groups [16], [19], [29], [32]. If the theory described above is correct, then one might justifiably wonder why these short-term measurements from various groups on the state-of-the-art samples and careful measurement setup provide a diverse set of initial exponents as well as various transition times to the steady-state $n \sim 1/6$ regime, not a universal curve anticipated in the R-D model discussed in Sections II-A and B. Indeed, whereas UFM done by Reisinger *et al.* [19] shows a steady state before $\sim$1 s, the same signature is observed at $\sim$1 s [16] and $\sim$100 s [29], [32] for OTFM measurements. This puzzle is resolved (discussed below) by a careful consideration of initial-delay contamination in the so-called "delay-free" OTFM experiments (similar delay contamination in OTFM is also reported in [28] and [37]).

Consider the definitions of ΔV_T in OTFM measurements: 1) $\Delta V_T(t) = |(I_{\rm dlin}(t) - I_{\rm dlin}(t_0))/I_{\rm dlin}(t_0)|^* V_{\rm GT0}$ [15], [16], [29], 2) $\Delta V_T(t) = M|(I_{\rm dlin}(t) - I_{\rm dlin}(t_0))/I_{\rm dlin}(t_0)|$ [30],

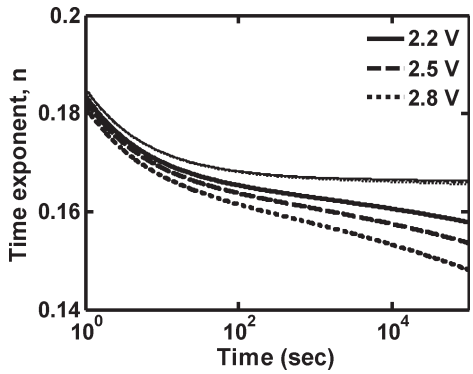


Fig. 5. Decrease in long-term NBTI exponent results from the reduction of stress field (for constant voltage stress) with time. Here, thick lines consider such reduction in electric field over time, whereas thin lines do not consider such "self-saturation" effect.

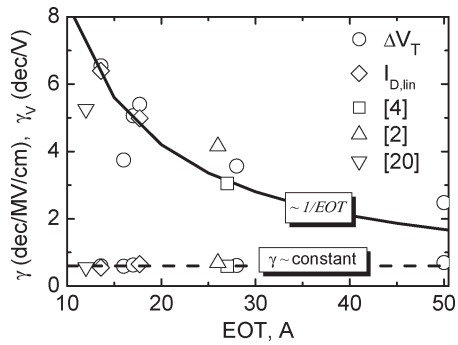


Fig. 6. γ and γ_V variation with EOT. As predicted for the field-dependent NBTI model, (dashed) γ remains almost constant, and (solid) γ_V keeps around $1/\text{EOT}$ line. For ΔV_T measurements, necessary delay corrections have been accounted for.

[32], etc. These methods give comparable results for long-term NBTI stress [30]. Obviously, an exact ΔV_T measurement using OTFM necessitates the use of $t_0 = 0$. Yet, OTFM setups (using conventional source-measure unit-based electronics) require some time (≥ 1 ms) before measuring $I_{\rm dlin}(t_0)$. Hence, $t_0 \neq 0$ (termed as *time-zero* delay) in such measurements, which will introduce an error in degradation estimates. Fig. 4(b) shows the solution of the R-D model with various t_0. Here, the UFM of [19] is considered to have $t_0 \sim 0$ because the first measurement of ΔV_T is done before the device is stressed; moreover, the $\sim$1-μs measurement delay used in the UFM shows a negligible effect [19, Fig. 3]. It is clear from Fig. 4(b) that such a time-zero delay results in an "apparent and artificial" increase of n mainly in short-term NBTI degradation and also in a shift in the time at which steady state is reached. Therefore, for type I devices, our results imply that once the t_0 delay is corrected for, the OTFM results would be consistent with each other and predict the same degradation as the UFM method. Furthermore, these t_0-corrected experimental data can then be interpreted by the H$-$H$_2$ model (Sections II-A and B). If such t_0 correction in OTFM is difficult (as may be the case in practice), Fig. 4(b) also suggests that a reasonable projection for NBTI lifetime can still be made, for example, if t_0 is kept below 1 ms and the projection is based on the data above 10 s. We emphasize that if this short-term region (higher time exponent region) is not excluded during reliability projection, the degradation exponent will be contaminated by sub 10-s time exponent and will result in unnecessarily pessimistic lifetime prediction.

D. Long-Term Reduction in the NBTI Exponent

Finally, we consider the soft saturation of ΔV_T at a long stress time. Fig. 5 shows measured time exponent as a function of time for various stress biases. Note that the time exponent reduces from the robust $n \sim 1/6$ value at a long stress time (e.g., $\delta n \sim 0.005/\text{decade}$), the reduction being higher for higher stress V_G. This reduction arises from the decrease in oxide electric field at a constant voltage stress due to the increase in ΔV_T over time. The figure predicts a decrease in n to $\sim$0.15 for the stress condition at 2.8 V, 125 °C in Fig. 2, and such exponent is consistently observed in experiments [15], [16], [18], and [19].

To summarize this section on time dependence, we have both analytically and numerically shown that the short, long, and very-long term characteristics of time exponents of NBTI degradations for type I devices can be consistently explained by the R-D model.

III. Field Dependence Within the R-D Model

A. Field- Versus Voltage-Dependent Degradation

In order to explain how NBTI degradation for type I devices varies with stress voltage, we first need to identify whether NBTI is a field-dependent or a voltage-dependent phenomenon [2]. Applying same stress voltage to bias a PMOS in inversion and NMOS in accumulation, Mahapatra *et al.* [39] have shown that PMOS device, having higher electric field due to flat-band voltage difference, degrades more compared to the NMOS device. They have also shown that both PMOS and NMOS equally degrade for the same surface electric field. Therefore, interface trap generation is expected to be a field-dependent phenomenon in these devices.

This field dependence of interface trap generation, hence, NBTI for $N_{\rm IT}$-dominated type I devices, has been considered using several empirical models, e.g., exponential model $\Delta V_T \propto \exp(\gamma E)$ [20], [40], power-law model $\Delta V_T \propto (E)^p$ [41], mixed model $\Delta V_T \propto E\exp(\gamma E)$ [2], [12], [16], [18], etc. Although a study in [41] shows the power-law model to give better results compared to the exponential model, their experimental data have saturating trends, a possible artifact due to recovery [16]. In the next section, a physically based field-dependent model is developed by considering both the dipolar effects of Si–H bond as well as field-dependent hole tunneling. It is shown that $E\exp(\gamma E)$ model (developed on a more physical basis [2], [18]) gives a consistent explanation of the field dependence in NBTI. Such model avoids the necessity of NBTI field dependence that purely comes from dipolar dissociation (as suggested in [14]) and anticipates that neither the power law nor the exponential model gives a proper estimation of NBTI lifetime; rather, they misinterpret low-voltage NBTI degradation by overestimating or underestimating the lifetime, respectively, as shown later.

B. Experimental Validation of the Phenomenological $E\exp(\gamma E)$ Model

Fig. 6 shows the field acceleration parameter γ and the voltage acceleration factor γ_V, which is defined as the number of decades change in lifetime for unit change in gate voltage [2], as functions of EOT from a wide range of experiments done by several groups (also compared with some of our results). Here, we use (10) (the origin of which is discussed in the next section) for interpreting the experimental results and found that as EOT changes, γ remains almost constant ($\sim 0.6 \pm 0.05$) as expected for a field-dependent degradation, whereas γ_V varies as $\sim 1/\text{EOT}$. To understand the γ_V versus EOT dependence (for constant γ), consider two oxides of same EOT being stressed at two stress voltages V_1 (electric field, E_1) and V_2 (E_2) to produce the same NBTI degradation after time t_1 and t_2, respectively, so that $E_1^{2/3}\exp(2\gamma E_1/3)t_1^n = E_2^{2/3}\exp(2\gamma E_2/3)t_2^n$ by (10). Using $E \sim V/\text{EOT}$ (where units of E, V, and EOT are MV/cm, V, and Å, respectively) and $n \sim 1/6$, we obtain

$$\gamma_V(\text{in dec/V}) \approx \frac{\log_{10}(t_1/t_2)}{(V_2 - V_1)} \sim \frac{40}{0.23(\text{EOT})}\left(\gamma + \frac{\ln(E_1/E_2)}{E_1 - E_2}\right). \quad (8)$$

Equation (8) suggests an inverse relationship between γ_V and EOT (see Fig. 6 and [2]) for the constant field acceleration parameter γ; however, the constant of proportionality depends on the electric field used to extract the data. This electric field dependence is clearly seen in the deviation from inverse EOT relationship in Fig. 6 due to the use of "different E_1 and E_2" for different devices. For example, $E_1 - E_2$ is higher for EOT $=$ 13 Å device compared to EOT $=$ 50 Å device. As a result, when $\sim 1/\text{EOT}$ line is drawn in Fig. 6 through EOT $=$ 13 Å data point, the drawn line falls below EOT $=$ 50 Å data point. As a result, γ_V values are on and above the solid line drawn in Fig. 6 for the respective devices considered.[3] Our analysis of degradation measurements performed by several groups validates the universality of the phenomenological $E\exp(\gamma E)$ field model. Moreover, we have used this model to fit the voltage-dependent degradation data of Fig. 2. Excellent fitting at all stress time also justifies the applicability of the model in explaining NBTI stress at any timescale.

C. Physical Hole-Assisted Field-Enhanced Thermal Dissociation Model

As shown in Section II [see the discussion below (7)], the "H$-$H$_2$ R-D" model anticipates that for type I devices at a long stress time,

$$N_{\text{IT}} = \left(\frac{k_H}{k_{H2}}\right)^{1/3}\left(\frac{k_f N_0}{k_r}\right)^{2/3}(6D_{H2}t)^{1/6}. \quad (9)$$

[3]As electric fields for devices (taken from references) considered in Fig. 6 are estimated, a distinction between $E\exp(\gamma E)$ and $\exp(\gamma E)$ models may appear to be tentative. However, based on the discussion in Section III-C and the prediction of deviation in Fig. 6 by (8), $E\exp(\gamma E)$ model appears more plausible.

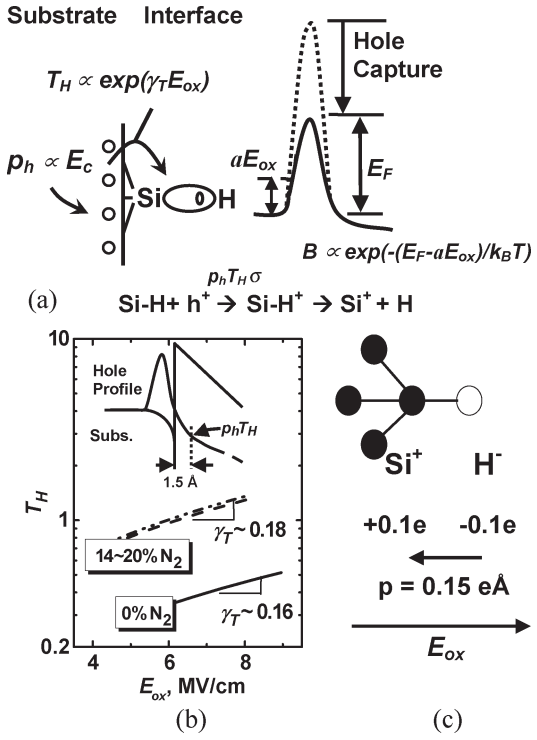


Fig. 7. (a) Model for hole-assisted field-enhanced thermal dissociation of Si–H bond resulting $k_f \propto E_c\exp(\gamma_T E_{\text{ox}})\exp[-(E_F - aE_{\text{ox}})/k_BT] \propto E_c\exp(\gamma E_{\text{ox}})$. Resultant field acceleration factor can be written as $\gamma = \gamma_T + a/k_BT$. (b) Hole penetration at 1.5 Å (for clarity, the band diagram in the inset is drawn upside-down), which is modeled using $A_1E_c\exp(\gamma_T E_{\text{ox}})$, gives an estimate of γ_T. (c) Polarization properties of Si–H bond (polarization vector is opposite in direction of the applied field under NBTI stress).

Now, the field dependence of N_{IT} generation must reflect the field dependencies of k_f, k_r, D_{H2}, k_H, and k_{H2}. As the diffusing hydrogen species is neutral [2], [18], k_r, D_{H2}, k_H, and k_{H2} are field independent. As such, the field dependence must be encapsulated in k_f and is explained by Fig. 7(a). The proposed model for k_f assumes that inversion layer holes with concentration p_h ($\propto E_c$; the electric field due to mobile carriers, which excludes the depletion charge contribution from total electric field E_{ox} in inversion and equals E_{ox} in accumulation) near the Si/dielectric interface tunnel into [with tunneling probability of $T_H \sim \exp(\gamma_T E_{\text{ox}})$, where γ_T is a constant that depends on tunneling environment; see Fig. 7(b)] and are captured by Si–H bonds (capture cross section, σ) $\sim$1.5 Å (Si–H bond length) away from the interface, which leads to a hole-assisted field-enhanced thermal generation of interface traps, having a reaction rate $B \propto \exp(-(E_F - aE_{\text{ox}})/k_BT)$ [see Fig. 7(a) and (c)]. Here, E_F represents the activation energy for Si–H bond dissociation (after a hole is captured),[4] and aE_{ox} represents thermal barrier lowering due to positioning

[4]Please note that although the first-principle calculation of E_F [42] questions the form of Si–H dissociation that is considered in the R-D model, it anticipates an activation energy of 0.36 eV, which is completely different from experimental observations of $\sim$0.1 eV [12], [15]–[19], [33]. However, the calculation in [42] suggests a reduction in Si–H dissociation energy with hole capture, which is used in our field model.

of the polar Si–H bond in the electric field, with a being the effective dipole moment [43]. In sum, $k_f \propto p_h T_H \sigma B$, so that

$$\Delta V_{\rm IT} = \frac{q\,(\alpha \Delta N_{\rm IT})}{C_{\rm ox}} = A^* {\rm EOT}^* (E_c)^{2/3} \exp\left(\frac{2\gamma E_{\rm ox}}{3}\right) \exp\left(-\frac{nE_{D1}}{k_B T}\right) t^n \tag{10}$$

where

$$\gamma = \gamma_T + a/k_B T \tag{11a}$$

$$nE_{D_1} = nE_D + 2/3(E_F - E_R). \tag{11b}$$

E_R and E_D are the activation energy for k_r and the combined activation energy for D_{H2}, k_H, and k_{H2}, respectively. For (10), $C_{\rm ox}$ is the dielectric capacitance, and EOT is obtained from a fit of experimental C–V using simulators like [44] that includes the effects of multisubband electron/hole quantization, polydepletion, etc. The overall activation energy for $\Delta V_{\rm IT}$ from (10) can also be written as

$$E_T \equiv nE_A = nE_D + 2/3(E_F - E_R - aE_{\rm ox}). \tag{11c}$$

D. Theoretical Estimates of γ_T and a

To theoretically estimate γ_T for (11a), first, we determine the amount of hole that can reach Si–H bond (∼1.5 Å away from the inverted channel) using the self-consistent Schrödinger–Poisson solution of MOS electrostatics [44]. Electric field dependence of hole penetration $(p_h T_H)$, when modeled using $A_1 E_c \exp(\gamma_T E_{\rm ox})$ [Fig. 7(b)], gives γ_T in the range of 0.16–0.18, which is approximately independent of the nitrogen concentration in the dielectric.

Next, the parameter a is estimated as follows [Fig. 7(c)]. According to the Pauling scale [45], electronegativity difference between Si $(x_{\rm Si} = 1.8)$ and H $(x_H = 2.1)$ is 0.3. The corresponding ionic bond energy $U_{H-{\rm Si}} = 1.3(x_{\rm Si} - x_H)^2 = 0.1$ eV allows one to estimate the effective charge transfer within Si–H dipole, i.e., $\sqrt{(4\pi\varepsilon_0 r U_{H-{\rm Si}}} = 0.1q$ [43], [45], where q is the electron charge, and r is the Si–H bond length (∼1.5 Å).[5] The resultant dipole moment in Si–H bond is $p = 0.15q$ Å. Considering that the Si–H dipole resembles a dipolar orientation normal to a thin slab, the local electric field [47] is given by $E_{\rm loc} = (1 + L\chi_{\rm int})E_{\rm ox} = \varepsilon_{\rm int} E_{\rm ox} = (3.9 - 8)E_{\rm ox}$, where $L = 1$ (depolarization factor [47]), $\chi_{\rm int}$ is the dielectric susceptibility at interface, and $\varepsilon_{\rm int}$ is the dielectric permittivity (relative) at interface having values approximately 3.9 (SiO_2) to 8 (average of SiO_2 and Si) [48]. Therefore, $a = p\varepsilon_{\rm int} = 0.6{-}1.2q$ Å. In other words, when the Si–H bond is placed in an electric field, applied opposite to the polarization vector (which is the field polarity for NBTI stress), its energy will be increased by $pE_{\rm loc} = p\varepsilon_{\rm int} E_{\rm ox} = aE_{\rm ox}$.

[5]Similar charge transfer is also calculated in [46] based on Sanderson's scale.

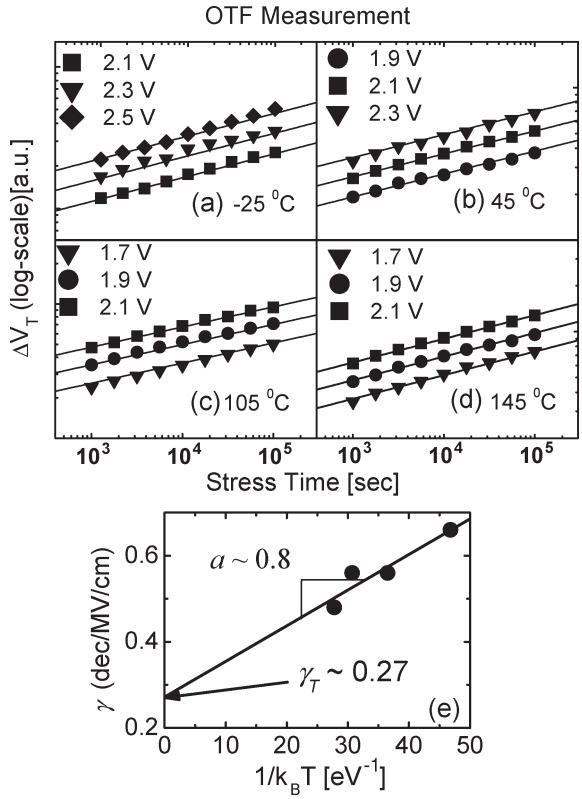


Fig. 8. (a–d) Fitting ΔV_T (measured by OTFM) versus t (log–log scale) at different T and V_G by (10) gives A, γ, and n as fitting parameters. Here, $n = 0.15{-}0.16$, and N_2 dose is 8%. (e) Obtained γ is plotted against $1/k_B T$, which results in $\gamma_T = 0.27 \pm 0.05$ decade/MV/cm and $a = 0.8 \pm 0.15q$ Å.

E. Experimental Determination of γ_T and a

In experimentally determining γ_T and a, first, we use (10) to fit a wide range of experimental ΔV_T versus t data that are measured at different voltages and temperatures, as shown in Fig. 8(a)–(d), which gives A, γ, and n as the fitting parameters for each temperature (effect of E_{D1} can be included in A). A plot of γ versus $1/k_B T$ gives the tunneling parameter γ_T as the intercept at $T \to \infty$ and the effective dipole moment a as the slope [see (11a)]. Alternately, as E_A changes with $E_{\rm ox}$ [see (11c)], $0.25 dE_A/dE_{\rm ox}$ can also be used to determine a [18]. A statistical summary of a, which is obtained using different experimental and theoretical approaches, is presented in Table I and is remarkably consistent with the values obtained by our theory and measurements.

F. Predicting Lifetime Using Field Model (for $N_{\rm IT}$-Specific Degradation)

The field-dependent model that is just developed can be used to calculate the safe operating voltage for a given lifetime (say, 5 years) and failure criteria (e.g., 60 mV of ΔV_T; Fig. 9). Our calculation suggests that neither the exponential nor the power-law fit of ΔV_T data (taken at higher stress condition) provides exact estimations of safe operating voltage. The exponential fit

TABLE I
COMPARING a DETERMINED BY VARIOUS METHODOLOGIES

a [qÅ]	Obtained from	
0.8	$\gamma = \gamma_T + a/k_BT$	Fig. 9b
0.98 (avg.)	$a = 0.25\ dE_A/dE_{ox}$	[18]
1.47	$a = (\gamma - \gamma_T)k_BT$ *	[14]
0.6~1.2	Si-H polarization	Sec. III (D)

* γ obtained after delay correction and using $\gamma_T \sim 0.27$ (from Fig. 8). In [14], $a = 3.2$ qÅ

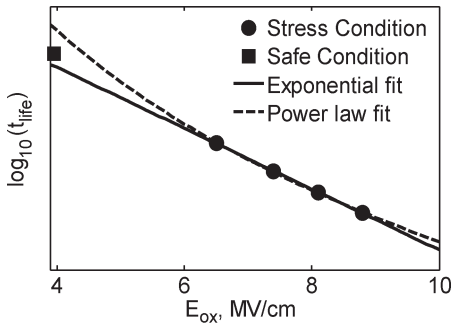


Fig. 9. Comparison between the exponential and the power-law fit of ΔV_T data for obtaining safe operating condition using stress data. The exponential fit underestimates lifetime, whereas the power-law fit overestimates.

underestimates the lifetime for the device considered, whereas the power-law fit overestimates. The power-law fit can provide a good estimate if low stress field data are used in lifetime projection (which will be time consuming).

In summary of this section, we observe that part of the field dependencies of $N_{\rm IT}$-specific NBTI in type I devices arises from $E_{\rm ox}$ dependence of T_H and another part from $E_{\rm ox}$-induced thermal barrier lowering. This, we believe, is the first consistent interpretation of field dependence of NBTI degradation within the R-D framework. Previous analysis of delay-contaminated NBTI data [14] suggested experimental values of a that were three to four times larger than theoretically expected values. It appears that similar to resolution of puzzles that involve time exponent $(n = 0.25 \rightarrow 0.16)$ and temperature dispersion parameter $(d = 1 - k_BT/E_0 = 0.6 \rightarrow 0$, where E_0 is the characteristic width for density of states of hydrogen-trapping sites in dispersive transport [6], [16]), zero-delay measurements and partitioning γ into γ_T and a/k_BT (previous work incorrectly assigns $\gamma = a/k_BT$ [14] or to $\gamma = \gamma_T$ [2] exclusively) hold the key for the consistent estimation of field acceleration of NBTI degradation.

IV. HOLE TRAPPING UNDER NBTI STRESS

The paper presented so far analyzes type I devices having SiO_2 and plasma SiON gate dielectric, where NBTI predominantly results from $N_{\rm IT}$ generation. There could be cases, however, where additional hole trapping [term 2 in (1)] is known to affect NBTI degradation, particularly in type II devices with highly nitrided dielectric [28], [32], [33] or in high-K materials [35]. For these films, additional contributions from hole trapping must be explicitly considered.

There are three points of view regarding the relative contribution of hole trapping to $N_{\rm IT}$ generation of NBTI [see (1)].

1) Some suggest that NBTI is primarily a phenomenon of (dispersive) hole trapping (for all devices), and any contribution from $\Delta N_{\rm IT}$, if present, is small [28], [30].
2) *Both* hole trapping and $\Delta N_{\rm IT}$ are present; however, $\Delta N_{\rm IT}$ involves (dispersive) reaction-limited generation of interface traps [5], [32].
3) Others believe that hole trapping saturates very quickly (approximately tens of milliseconds), and any subsequent shift in ΔV_T reflects diffusion-limited generation of interface traps within the R-D framework [19].[6]

As hole trapping has negligible temperature activation, viewpoint 1 cannot explain experimental results for type I devices [33], although it may be the case for type II devices [33]. Note that the original presumption that even type I devices are dominated by hole trapping is based on the erroneous comparison of charge pumping with I_D–V_G (or OTFM) stress data, without considering $N_{\rm IT}$ relaxation due to measurement delay and difference in scanned energy zones (within the bandgap) between the two technique, as discussed in detail in [33]. Viewpoint 2 suggests T-dependent long-term power-law time exponent [5], [32], which has been shown to be an artifact of measurement delay [16]. Accordingly, we will later generalize the third viewpoint to interpret the experimental results in Section IV-B.

A. Hole Trapping in Oxides With High Interfacial N Concentration

To investigate the effect of hole trapping in type II devices, we use a simple *elastic* trapping/detrapping model, in which holes from the inversion layer are first trapped in the preexisting defects within the oxide, and then, these trapped holes detrap toward the empty states in poly-Si and channel. Trap filling probability f_T evolves with time according to

$$\frac{df_T}{dt} = \sigma v_{\rm th}\left[p_h T_1(1 - f_T) - n_S T_1 f_T - n_G T_2 f_T\right] \qquad (12)$$

where p_h is the inversion layer hole density, $v_{\rm th}$ is the thermal velocity, n_S and n_G are the concentration of detrapping states at substrate and poly-Si, respectively, and T_1 and T_2 are the tunneling probability (obtained using the Wentzel–Kramers–Brillouin approximation) of holes from Si/dielectric interface to trap and trap to poly, respectively. Note that $\rho_H(x, E_i, t) = N_0\delta(E - E_i)f_T(x, E_i, t)$ for (1), where N_0 is the density of preexisting traps, and E_i is the energy level at which *elastic* trapping is possible. Let us emphasize

[6]Note that for the specific case of [19], the discussion of Fig. 2 suggests that any contributions to ΔV_T from hole trapping for this particular experiment are actually negligible, and the experimental data 125 °C can solely be understood in terms of the generalized R-D model. In fact, since activation energies for hole trapping, H and H_2 diffusion, and $N_{\rm IT}$ generation [12], [16], [49], [50] (represented by $E_{A({\rm HT})}$, $E_{A(H)}$, $E_{A(H2)}$, $E_{A({\rm IT})}$, respectively) are such that $E_{A({\rm HT})} \ll E_{A(H)} < E_{A(H_2)} \sim E_{A({\rm IT})}$, therefore, models that involve hole trapping would be inconsistent with temperature-dependent data for type-I devices.

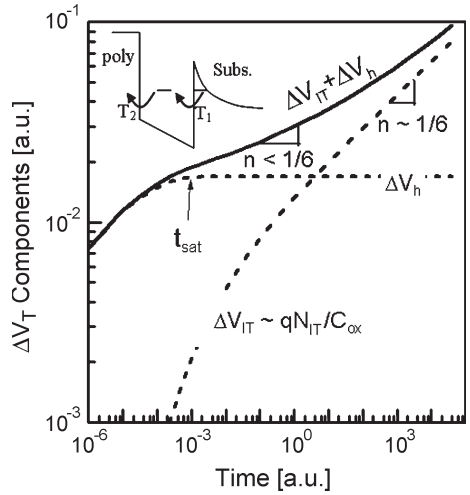


Fig. 10. Simulated time evolution of $\Delta V_{\rm IT}$, ΔV_H, and total ΔV_T for device, having thick oxide with uniform distribution of traps. Only elastic hole trapping is considered in the calculation of ΔV_H. As a result, hole trapping controls total degradation at a short stress time. It also results in a decrease in n at an intermediate stress time. Inset shows tunneling probabilities T_1 (into/out of traps from/toward substrate) and T_2 (into/out of traps from/toward gate) for the elastic hole trapping/detrapping model that is considered in the text.

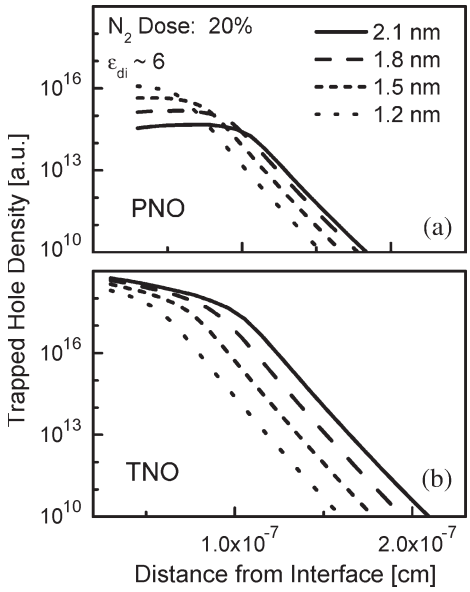


Fig. 11. Trapped hole profile for (a) plasma and (b) thermal nitrided oxides, having similar $\%N_2$ dose. Thermal-nitrided oxides trap more holes compared to plasma-nitrided oxides. Only elastic traps are considered in simulation.

that the role of detrapping—although negligible in thick films (as for the case in [51])—is fundamentally important in the trapping dynamics of thin films. The detrapping process in thin oxides limits the possibility of hole trapping in sites located near the gate—an important consideration that remains poorly appreciated in NBTI literature.

Analytical solution of (12) gives

$$f_T = \frac{T_1\left[1-\exp\left(-\sigma v_{\rm th}\left(p_h T_1 + n_s T_1 + n_G T_2\right)t\right)\right]}{\left(1+n_S/p_h\right)T_1 + n_G T_2/p_h}. \quad (13)$$

We estimate the maximum occupancy $f_{T(\max)} = T_1/[(1+n_S/p_h)T_1 + n_G T_2/p_h]$ at different trapping sites, which involves elastic tunneling of carriers into and out of traps (named as *elastic traps*) with tunneling probability T_1 for trap/substrate transition and T_2 for trap/gate transition (inset of Fig. 10), for a nitrided sample. Here, we assume $\sigma = 10^{-17}$ cm^2 (which is within the range of $\sim 10^{-13}$ cm^2 [52] and $\sim 10^{-20}$ cm^2 [35] used in literature for bulk trapping), $p_h = 10^{20}$ cm^{-3} at 4 eV above valence band tip at the Si/dielectric interface (typical value in an inverted channel, verified using [44]), $n_S \sim 0$ (no empty states in substrate, already occupied by holes), and $n_G = 10^{20}$ cm^{-3} (approximate electron concentration in the valence band of poly within an energy range of $\sim k_B T$, which can act as elastic hole detrapping sites). Using $T_{\rm PHY} = 37$ Å (corresponding to EOT = 24 Å for nitrided oxide having $\sim$25% N_2 dose [18]), calculated $f_{T(\max)}$ for hole traps that are located at 20 Å is approximately four orders-of-magnitude lower compared to that for the site that is located at 10 Å. Hence, the time constant $\tau = 1/\sigma v_{\rm th}(p_h T_1 + n_s T_1 + n_G T_2)$ for the hole trapping component of $\Delta V_T(\Delta V_H)$ will be dominated by the traps that are located close to the substrate. Additional contributions due to *inelastic* capture in shallow level hole traps may broaden the trapping time constant. Finally, inelastic trapping into deep level hole traps (discussed in [34] and [53]) may further increase the time constant. Moreover, if detrapping from deep traps is blocked by poly-bandgap, contributions from the traps close to gate may also become possible.

B. Verification of the Hole Trapping Model

We now verify the hypothesis of hole trapping discussed above by systematically exploring the thickness-dependent time exponents (within the range of $1-10^3$ s) for oxides with low (plasma nitrided [13], [54]) and high (thermal nitrided [13]) interfacial nitrogen (details are discussed in [33]). We assume that the hole traps are proportionally distributed to (presumed) N-profile [cf. Fig. 1(a)] in space. The simple model involving the elastic hole traps results in the saturation of ΔV_H within $\sim$1 ms (Fig. 10). Thus, overall ΔV_T shows $n < 1/6$ due to an additional contribution from ΔV_H at the timescale of $\sim$1–1000 s. Our calculations show that *both* plasma-nitrided oxide [PNO, cf. Fig. 11(a)] and thermal-nitrided oxide [TNO, cf. Fig. 11(b)] films have contributions from hole trapping. The profile decays from Si/oxide interface toward oxide/poly-Si interface because of hole detrapping into empty states of poly. For thicker oxides, the number of trapped holes increases because detrapping to empty states in poly is difficult. Moreover, since the centroids of the trapped charges shift away from the poly/oxide interface, in general, thicker oxides have larger contributions to ΔV_T from hole trapping. However, the N-profile in the PNO films, with a peak near the oxide/poly-Si interface [Fig. 1(a)], makes the contribution from hole trapping in PNO films approximately two orders-of-magnitude smaller than those of TNO films, with a peak near the oxide/substrate interface [Fig. 1(a)]. As such, the time exponents of PNO continue to be dominated by $\Delta V_{\rm IT}$ and remain essentially unchanged with thickness (see Fig. 12). However, the contributions from hole trapping in TNO film are significant enough

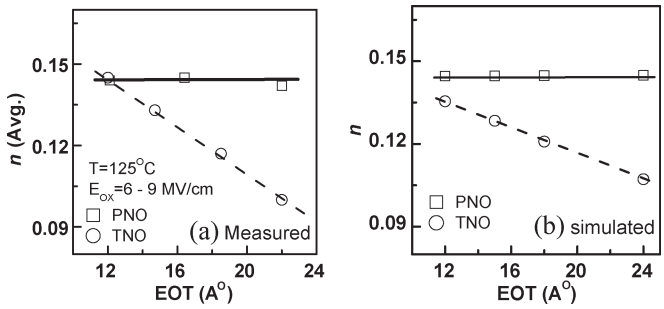


Fig. 12. (a) Experimental thickness dependence of average power-law time exponents obtained from on-the-fly $I_{\rm dlin}$ in PNO and TNO devices. (b) Simulated (involving elastic hole traps) thickness dependence under the same condition.

[Fig. 11(b)] to reduce the overall ΔV_T time exponent (Fig. 12), and one anticipates a reduction in n with increasing oxide thickness. In short, the role of hole trapping in the devices studied is significant for any thick film with significant interfacial N concentration and must be accounted for as an additional contribution to overall NBTI degradation. Therefore, $n < 0.1$ reported in [28] can be attributed to the existence of increased amount of traps within the oxide, in addition to the suppression of relative contribution from interface traps to ΔV_T in room temperature. Finally, we believe that the relative contributions of deep traps in the samples that are considered in this study are small because the model based on simple elastic trapping/detrapping appears to explain the experimental data (cf. Fig. 12) well. This does not exclude the possibility of such deep trapping in other films [28], [32], [34].

V. NBTI Relaxation

The discussion in Section II-D focused on the stress phase of NBTI degradation due to $N_{\rm IT}$ generation and hole trapping, as per (1). Once the effect of measurement delay is corrected for by using the OTFM or UFM techniques, the NBTI results are generally understood in terms of the R-D model and the hole trapping kinetics in type I and type II films.

NBTI relaxation, however, is an important consideration, particularly for typical CMOS operation where input voltage switches between "0" and "1" repeatedly. The R-D model for $N_{\rm IT}$ relaxation anticipates a relaxation of the form $N_{\rm IT} \sim N_{\rm IT}^{(0)}(1-\sqrt{t/2t_0}/\sqrt{1+t/t_0})$, where $N_{\rm IT}^{(0)}$ is the interface trap generated at $t = t_0$, that had appeared to be generally consistent with experiment [24]. Recently, however, there are new reports of *logarithmic* time ($\log t$) dependence of NBTI recovery based on the UFM for type I devices stressed at 125 °C [19] and the ultrafast OTFM for type II devices stressed at room temperature [28]. The NBTI relaxation by UFM spans over a timescale of eight to ten orders of magnitude, whereas the R-D model anticipates a slower recovery of over 1–2 decades in timescale [19], [31]. Although the fast relaxation in type II devices [28], [34] is understood as a consequence of hole detrapping and, as such, has not been particularly surprising, however, the fast and sustained $\log t$ relaxation in long-term 125 °C stressed type I films [19] is more difficult to understand because the stress phase of type I films appears to be dominated by $N_{\rm IT}$ generation with negligible contribution from hole trapping at high temperature (discussed in Sections II and IV). Therefore, this substantial discrepancy in interpreting the recovery experiments in [19] has instigated researchers to look for alternate models that involve hole detrapping [19], [28], dispersive diffusion [6], etc., in explaining NBTI recovery *even for type I devices*.

Here, we outline a model that explains this $\log t$ dependence of NBTI recovery for type I films and that fits naturally within the R-D framework (details will be discussed elsewhere). We attribute the discrepancy between experiment and theory to the inappropriate use of the quasi steady-state relationship $\Delta V_T = q\Delta N_{\rm IT}/C_{\rm ox}$ to ultrafast transient conditions. For example, during the relaxation phase, the capture of valence band electron by donor-type interface traps ($\mathrm{Si}^+ + \mathrm{e}^- \rightarrow \mathrm{Si}$) neutralizes the donor and is reflected in the *measured* $\log t$ dependence of NBTI recovery of ΔV_T; however, this neutralization does not cause $N_{\rm IT}$ recovery as calculated by the R-D model. $N_{\rm IT}$ recovery occurs only through repassivation by hydrogen species that diffuse back toward the interface—a much slower process. In sum, dynamics of electron capture/neutralization of donor states in the relaxation phase can initially result in $\log t$ dependence of measured ΔV_T recovery (as in [19]), followed by the slower $N_{\rm IT}$ recovery (as encapsulated by the R-D model) by back-diffusion of hydrogen species. As such, the perceived failure of the R-D model in explaining recovery experiments for type I devices, therefore, does not reflect an intrinsic limitation of the model itself; rather, it is a result of an improper interpretation of measured recovery using the R-D model. Details about the UFM-based NBTI recovery within the R-D framework will be published elsewhere.

VI. Conclusion

In this paper, we have systematically explored the time-dependent NBTI degradation over all timescales for different types of devices that have wide variation in nitrogen concentration within the dielectric film. Our analyses not only provide a sound physical basis of NBTI degradation but also allow us to resolve differences in NBTI measurements based on the OTFM and UFM methods. We have also developed and validated the first physical explanation of field dependence within the R-D framework, describing NBTI experiments for devices where ΔV_T is dominated by $N_{\rm IT}$ generation. Analyses that involve thick oxide devices with high nitrogen profile at the interface. Highlight the importance of hole trapping in determining the characteristics of NBTI degradation. Finally, we have briefly discussed ultrafast NBTI relaxation within the R-D framework, which we believe will be able to explain the fast transient recovery of NBTI. Our discussion broadens the theoretical framework of NBTI degradation in such a way that it may help the conceptual organization of existing experiments and design of newer experiments to further refine the NBTI analysis.

Acknowledgment

We acknowledge Dr. A. T. Krishnan for providing us the data for Fig. 8 and for other valuable discussions. We also

acknowledge Network of Computational Nanotechnology for providing computational resources.

REFERENCES

[1] B. E. Deal, M. Sklar, A. S. Grove, and E. H. Snow, "Characteristics of surface-state charge (Qss) of thermally oxidized silicon," *J. Electrochem. Soc.*, vol. 114, no. 3, pp. 266–274, Mar. 1967.

[2] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectron. Reliab.*, vol. 45, no. 1, pp. 71–81, Jan. 2005.

[3] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, Jul. 2003.

[4] S. Chakravarthi, A. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, "A comprehensive framework for predictive modeling of negative bias temperature instability," in *Proc. IEEE IRPS*, 2004, pp. 273–282.

[5] V. Huard, M. Denais, F. Perrier, N. Revil, C. Parthasarathy, A. Bravaix, and E. Vincent, "A thorough investigation of MOSFETs NBTI degradation," *Microelectron. Reliab.*, vol. 45, no. 1, pp. 83–98, Jan. 2005.

[6] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken, and M. Goodwin, "Disorder-controlled-kinetics model for negative bias temperature instability and its experimental verification," in *Proc. IEEE IRPS*, 2005, pp. 381–387.

[7] M. A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, "A comprehensive model for PMOS NBTI degradation: Recent progress," *Microelectron. Reliab.*, vol. 47, no. 6, pp. 853–862, Jun. 2007.

[8] *International Technology Roadmap for Semiconductors*, 2005.

[9] S. S. Tan, T. P. Chen, J. M. Soon, K. P. Loh, C. H. Ang, and L. Chan, "Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations," *Appl. Phys. Lett.*, vol. 82, no. 12, pp. 1881–1883, Mar. 2003.

[10] Y. Mitani, "Influence of nitrogen in ultra-thin SiON on negative bias temperature instability under AC stress," in *IEDM Tech. Dig.*, 2004, pp. 117–120.

[11] M. L. Green, E. P. Gusev, R. Degraeve, and E. L. Garfunkel, "Ultrathin (< 4 nm) SiO2 and Si-O-N gate dielectric layers for silicon microelectronics: Understanding the processing, structure, and physical and electrical limits," *J. Appl. Phys.*, vol. 90, no. 5, pp. 2057–2121, Sep. 2001.

[12] A. T. Krishnan, S. Chakravarthi, P. Nicollian, V. Reddy, and S. Krishnan, "Negative bias temperature instability mechanism: The role of molecular hydrogen," *Appl. Phys. Lett.*, vol. 88, no. 15, p. 153 518, Apr. 2006.

[13] J. R. Shallenberger, D. A. Cole, and S. W. Novak, "Characterization of silicon oxynitride thin films by X-ray photoelectron spectroscopy," *J. Vac. Sci. Technol. A, Vac. Surf. Films*, vol. 17, no. 4, pp. 1086–1090, Jul. 1999.

[14] K. O. Jeppson and C. M. Svensson, "Negative bias stress of mos devices at high electric-fields and degradation of Mnos devices," *J. Appl. Phys.*, vol. 48, no. 5, pp. 2004–2014, May 1977.

[15] G. Gupta, S. Mahapatra, L. Madhav, D. Varghese, K. Ahmed, and F. Nouri, "Interface-trap driven NBTI for ultrathin (EOT $\sim$ 12Å) plasma and thermal nitrided oxynitrides," in *Proc. IEEE IRPS*, 2006, pp. 731–732.

[16] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. A. Alam, "On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications," in *IEDM Tech. Dig.*, 2005, pp. 684–687.

[17] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, S. Krishnan, and L. Levitov, "Material dependence of hydrogen diffusion: Implications for NBTI degradation," in *IEDM Tech. Dig.*, 2005, pp. 688–691.

[18] A. E. Islam, G. Gupta, S. Mahapatra, A. Krishnan, K. Ahmed, F. Nouri, A. Oates, and M. A. Alam, "Gate leakage vs. NBTI in plasma nitrided oxides: Characterization, physical principles, and optimization," in *IEDM Tech. Dig.*, 2006, p. 12.4.1.

[19] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, "Analysis of NBTI degradation- and recovery- behaviour based on ultra fast V_T measurements," in *Proc. IEEE IRPS*, 2006, pp. 448–453.

[20] C. L. Chen, Y. M. Lin, C. J. Wang, and K. Wu, "A new finding on NBTI lifetime model and an investigation on NBTI degradation characteristic for 1.2 nm ultra thin oxide," in *Proc. IEEE IRPS*, 2005, pp. 704–705.

[21] N. K. Jha, P. S. Reddy, and V. R. Rao, "A new drain voltage enhanced NBTI degradation mechanism," in *Proc. IEEE IRPS*, 2005, pp. 524–528.

[22] S. Zafar, "Statistical mechanics based model for negative bias temperature instability induced degradation," *J. Appl. Phys.*, vol. 97, no. 10, p. 103 709, May 2005.

[23] J. B. Yang, T. P. Chen, S. S. Tan, and L. Chan, "Analytical reaction-diffusion model and the modeling of nitrogen-enhanced negative bias temperature instability," *Appl. Phys. Lett.*, vol. 88, no. 17, p. 172 109, Apr. 2006.

[24] M. A. Alam, "A critical examination of the mechanics of dynamic NBTI for PMOSFETs," in *IEDM Tech. Dig.*, 2003, pp. 345–348.

[25] S. Kumar, C. H. Kim, and S. S. Sapatnekar, "An analytical model for negative bias temperature instability," in *Proc. ICCAD*, 2006, p. 6D.1.

[26] G. Chen, M. F. Li, C. H. Ang, J. Z. Zheng, and D. L. Kwong, "Dynamic NBTI of p-MOS transistors and its impact on MOSFET scaling," *IEEE Electron Device Lett.*, vol. 23, no. 12, pp. 734–736, Dec. 2002.

[27] R. Fernández, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodríguez, M. Nafría, and G. Groeseneken, "AC NBTI studied in the 1 Hz–2 GHz range on dedicated on-chip CMOS circuits," in *IEDM Tech. Dig.*, 2006, p. 12.6.1.

[28] C. Shen, M. F. Li, C. E. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y. C. Yeo, "Characterization and physical origin of fast Vth transient in NBTI of pMOSFETs with SiON dielectric," in *IEDM Tech. Dig.*, 2006, p. 12.5.1.

[29] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability [PMOSFETs]," in *IEDM Tech. Dig.*, 2003, pp. 341–344.

[30] C. R. Parthasarathy, M. Denais, V. Huard, G. Ribes, E. Vincent, and A. Bravaix, "New insights into recovery characteristics post NBTI stress," in *Proc. IEEE IRPS*, 2006, pp. 471–477.

[31] T. Grasser, W. Gos, V. Sverdlov, and B. Kaczer, "The universality of NBTI relaxation and its implications for modeling and characterization," in *Proc. IEEE IRPS*, 2007, pp. 268–280.

[32] V. Huard, C. R. Parthasarathy, C. Guerin, and M. Denais, "Physical modeling of negative bias temperature instabilities for predictive extrapolation," in *Proc. IEEE IRPS*, 2006, pp. 733–734.

[33] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, M. A. Alam, "On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy?" in *Proc. IEEE IRPS*, 2007, pp. 1–9.

[34] D. S. Ang and S. Wang, "Recovery of the NBTI-stressed ultrathin gate p-MOSFET: The role of deep-level hole traps," *IEEE Electron Device Lett.*, vol. 27, no. 11, pp. 914–916, Nov. 2006.

[35] S. Zafar, A. Kumar, E. Gusev, and E. Cartier, "Threshold voltage instabilities in high-k gate dielectric stacks," *IEEE Trans. Device Mater. Rel.*, vol. 5, no. 1, pp. 45–64, Mar. 2005.

[36] H. Kufluoglu and M. A. Alam, "A generalized reaction-diffusion model with explicit $H-H_2$ dynamics for negative bias temperature instability (NBTI) degradation," *IEEE Trans. Electron Devices*, vol. 54, no. 5, pp. 1101–1107, May 2007.

[37] A. E. Islam, H. Kufluoglu, D. Varghese, and M. A. Alam, "Critical analysis of short-term negative bias temperature instability measurements: Explaining the effect of time-zero delay for on-the-fly measurements," *Appl. Phys. Lett.*, vol. 90, no. 8, p. 083 505, Feb. 2007.

[38] C. G. Van de Walle, "Theory of Hydrogen-related levels in semiconductors and oxides," in *IEDM Tech. Dig.*, 2005, pp. 400–403.

[39] S. Mahapatra, P. B. Kumar, and M. A. Alam, "A new observation of enhanced bias temperature instability in thin gate oxide p-MOSFET," in *IEDM Tech. Dig.*, 2003, pp. 337–340.

[40] N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, and T. Horiuchi, "The impact of bias temperature instability for direct-tunneling ultra-thin gate oxide on MOSFET scaling," in *VLSI Symp. Tech. Dig.*, 1999, pp. 73–74.

[41] H. Aono, E. Murakami, K. Okuyama, A. Nishida, M. Minami, Y. Ooji, and K. Kubota, "Modeling of NBTI degradation and its impact on electric field dependence of the lifetime," in *Proc. IEEE IRPS*, 2004, pp. 23–27.

[42] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature instability," *Appl. Phys. Lett.*, vol. 86, no. 14, p. 142 103, Apr. 2005.

[43] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO2 thin films," *J. Appl. Phys.*, vol. 84, no. 3, pp. 1513–1523, Aug. 1998.

[44] A. Ghetti, A. Hamad, P. J. Silverman, H. Vaidya, and N. Zhao, "Self-consistent simulation of quantization effects and tunneling current in ultra-thin gate oxide MOS devices," in *Proc. SISPAD*, 1999, pp. 239–242.

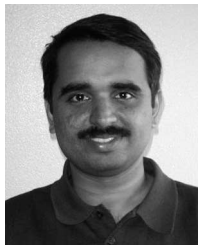
[45] L. Pauling, *The Nature of The Chemical Bond*, 3rd ed. Ithaca, NY: Cornell Univ. Press, 1960, pp. 88–91.
[46] H. Z. Massoud, "Charge-transfer dipole-moments at the Si-SiO_2 interface," *J. Appl. Phys.*, vol. 63, no. 6, pp. 2000–2005, Mar. 1988.
[47] C. Kittel, *Introduction to Solid State Physics*, 7th ed. Hoboken, NJ: Wiley, 1996, pp. 381–390.
[48] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, "The electronic structure at the atomic scale of ultrathin gate oxides," *Nature*, vol. 399, no. 6738, pp. 758–761, Jun. 1999.
[49] B. J. Fishbein, J. T. Watt, and J. D. Plummer, "Time resolved annealing of interface traps in polysilicon gate metal-oxide-silicon capacitors," *J. Electrochem. Soc.*, vol. 134, no. 3, pp. 674–681, Mar. 1987.
[50] B. Tuttle, "Energetics and diffusion of hydrogen in SiO2," *Phys. Rev. B, Condens. Matter*, vol. 61, no. 7, pp. 4417–4420, Feb. 2000.
[51] L. Lundkvist, I. Lundstrom, and C. Svensson, "Discharge of MNOS structures," *Solid State Electron.*, vol. 16, no. 7, pp. 811–818, 1973.
[52] T. H. Ning, "Capture cross-section and trap concentration of holes in silicon dioxide," *J. Appl. Phys.*, vol. 47, no. 3, pp. 1079–1081, Mar. 1976.
[53] J. F. Zhang, C. Z. Zhao, A. H. Chen, G. D. Groeseneken, and R. Degraeve, "Hole traps in silicon dioxides—Part I: Properties," *IEEE Trans. Electron Devices*, vol. 51, no. 8, pp. 1267–1273, Aug. 2004.
[54] S. Rauf, S. Lim, and P. L. G. Ventzek, "Model for nitridation of nanoscale SiO2 thin films in pulsed inductively coupled N-2 plasma," *J. Appl. Phys.*, vol. 98, no. 2, p. 24 305, Jul. 2005.

Ahmad Ehteshamul Islam (S'02) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2004. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.

During 2004–2005, he was a Lecturer with the Department of Electrical and Electronic Engineering, BUET. His research interests include modeling, simulation, characterization, and reliability of nanoscale devices. He is currently working with NBTI issues in devices having oxinitride and high-*K* gate dielectrics.

Haldun Kufluoglu (S'04) was born in Turkey in 1979. He received the B.S. and M.S. degrees in electrical engineering from Purdue University, West Lafayette, IN, where he is currently working toward the Ph.D. degree, in 2001 and 2003, respectively. Particularly, his Ph.D. research involves measurements and theoretical modeling of MOSFET degradation mechanisms such as NBTI, HCI, and TDDB, and their implications on VLSI design.

His research interests include MOSFET reliability, experimental characterization, and modeling of semiconductor devices. He also participates in OFF-state transistor reliability assessment. Previously, he obtained microfabrication skills in a MEMS-based sensor that was interfaced with live neurons for biological applications.



Dhanoop Varghese (S'06) received the B.Tech. degree in electronics and communication engineering from Regional Engineering College, Calicut, India, in 2002 and the M.Tech. degree in electrical engineering from Indian Institute of Technology, Bombay, India, in 2005. He is currently working toward the Ph.D. degree at Purdue University, West Lafayette, IN.

His current research interests are in the field of semiconductor device physics, simulation, modeling, and characterization. He has worked on bias temperature and hot carrier reliability issues in MOSFETs and high-*k* gate dielectrics.

Souvik Mahapatra (S'99–M'99) received the Ph.D. degree in electrical engineering from Indian Institute of Technology (IIT), Bombay, India, in 1999.

He was with Bell Laboratories, Murray Hill, NJ, during 2000 and 2001. Since 2002, he has been with the Department of Electrical Engineering, IIT, where he is currently an Associate Professor. His research interests are in the area of characterization, modeling and simulation of CMOS and Flash memory devices, and device reliability. He has published more than 60 papers in peer-reviewed journals and conferences, delivered invited talks in leading international conferences including the IEEE International Electron Devices Meeting, was a tutorial speaker at IEEE International Reliability Physics Symposium, and acted as a reviewer of several international journals and conferences.

Muhammad Ashraful Alam (F'05) received the B.S.E.E. degree from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 1988, the M.S. degree from Clarkson University, Potsdam, NY, in 1991, and the Ph.D. degree from Purdue University, Lafayette, IN, in 1994, all in electrical engineering.

He is a Professor of electrical and computer engineering at Purdue University, West Lafayette, IN, where his research and teaching focus on physics, simulation, characterization, and technology of classical and novel semiconductor devices, including theory of oxide reliability, nanocomposite thin film transistors, and nanobio sensors. From 1995 to 2001, he was with Bell Laboratories–Lucent Technologies, Murray Hill, NJ, as a member of the Technical Staff in the Silicon ULSI Research Department. From 2001 to 2003, he was a Distinguished Member of Technical Staff at Agere Systems, Murray Hill. He joined Purdue University in 2004. He has published over 80 papers in international journals and has presented many invited and contributed talks at international conferences.

Dr. Alam is an IEEE Fellow and was the recipient of the 2006 IEEE Kiyo Tomiyasu Award for contributions to device technology for communication systems.