

A comprehensive model of PMOS NBTI degradation

M.A. Alam^{a,*}, S. Mahapatra^b

^a Agere Systems, Allentown, PA, USA

^b Department of Electrical Engineering, IIT, Bombay, India

Received 13 November 2003; received in revised form 31 March 2004

Available online 3 August 2004

Abstract

Negative bias temperature instability has become an important reliability concern for ultra-scaled Silicon IC technology with significant implications for both analog and digital circuit design. In this paper, we construct a comprehensive model for NBTI phenomena within the framework of the standard reaction–diffusion model. We demonstrate how to solve the reaction–diffusion equations in a way that emphasizes the physical aspects of the degradation process and allows easy generalization of the existing work. We also augment this basic reaction–diffusion model by including the temperature and field-dependence of the NBTI phenomena so that reliability projections can be made under arbitrary circuit operating conditions.

© 2004 Published by Elsevier Ltd.

1. Introduction

1.1. Background

The instability of PMOS transistor parameters (e.g., threshold voltage, transconductance, saturation current, etc.) under negative (inversion) bias and relatively high temperature has been known to be a reliability concern since the 1970s [1–4], and modeling effort to understand this is also just as old [5]. Named negative bias temperature instability (NBTI) in analogy to the Na⁺ induced bias temperature instability (BTI) that caused similar shift in threshold voltages under operating conditions, the original technological motivation for NBTI studies was the shift of threshold voltage of p-channel MNOS [1,2] or FAMOS non-volatile memories [3] under high program and erase fields. New NMOS based non-volatile memory architectures eventually replaced

the PMOS based systems [6], and thereby reduced the importance of NBTI for those specific systems. However other processing and scaling changes, introduced over the last 30 years to improve device and circuit performances, have inadvertently reintroduced NBTI as a major reliability concern for mainstream analog and digital circuits [7–17]. These changes include:

- (a) Introduction of CMOS in early 1980s that has made PMOS and NMOS devices equally important for IC designs.
- (b) Introduction of dual poly-process that has allowed replacement of buried channel PMOS devices with surface channel PMOS devices. Although the circuit performance of surface channel device is better than that of buried channel device, their NBTI performance is actually worse [18,19].
- (c) Slower scaling of operating voltages for both analog and digital circuits compared to more aggressive scaling of oxide thickness has gradually increased the effective field across the oxide, which in turn has enhanced NBTI. Also, at a given oxide field, thin oxide devices were found to be more susceptible to NBTI than their thick oxide counterparts.
- (d) Thinner oxides have brought the poly-silicon gate closer the Si/SiO₂ interface. Note that diffusion of

* Corresponding author. Address: Room 313D, Electrical Engineering Building, School of Electrical and Computer Engineering, Purdue University, 465 Northwestern Avenue, West Lafayette, IN 47907-2035, USA. Tel.: +1-765-494-5988; fax: +1-765-494-6441.

E-mail address: alam@ecn.purdue.edu (M.A. Alam).

hydrogen away from the Si/SiO₂ interface controls NBTI-specific interface trap generation at the Si/SiO₂ interface. Since hydrogen diffusion through poly-silicon is faster than that in oxide, scaling of gate oxides has increased NBTI susceptibilities [16].

- (e) The introduction of nitrogen to reduce gate leakage and to inhibit boron penetration in thin oxide devices has made NBTI worse [10].

These technological factors, along with other circuit specific usages, have made NBTI as one of the foremost reliability concerns for modern integrated circuits (almost as important as of gate oxide TDDDB reliability). It is in this modern context that we need to understand the origin of the NBTI phenomena and the means to control it.

1.2. Experimental signatures of NBTI

Any theory of NBTI must be able to explain the following observations regarding the NBTI phenomena [7–27].

- (a) Time evolution is described by a power-law relationship (i.e., $\Delta V_T \sim t^n$ with $n \sim 1/4$) that holds over many decades of the timescale, ΔV_T being the threshold voltage shift, often attributed entirely to generation of interface traps, N_{IT} (more on this at the end of Section 2.1). However, there are intriguing deviations from the perfect, single-exponent power-law degradation that needs careful analysis. These deviations are mentioned below.
- (i) Increase in stress temperature increases n at early stress times. At higher temperatures, n becomes time-dependent and the degradation rate begins to saturate at longer times (e.g., $n \sim 0.25$ – 0.3 initially, but changes to $n \sim 0.2$ – 0.24 at later stage of degradation) [15,27].
 - (ii) Under stronger gate biases and especially for thick oxides, ΔV_T increases dramatically at later part of the stress (i.e., n changes from ~ 0.25 to ~ 0.5 as shown in Fig. 7). The exact point of this dramatic transition is temperature, bias and oxide thickness dependent [15].
 - (iii) For thinner oxides, NBTI degradation increases unexpectedly at later time at elevated stress temperatures [16].
- (b) A fraction of NBTI defects can be annealed once the stress is removed. This makes NBTI lifetimes (to reach a certain amount of degradation) higher for AC stress when compared to DC stress [20–23].
- (c) BTI appears to be associated with PMOS devices under inversion bias condition. However, NMOS devices at the same voltage show much lower NBTI [12].

- (d) Unlike voltage-dependent TDDDB phenomena [28], NBTI appears to be electric field dependent [1,2,5,15,16,25].
- (e) NBTI is an activated process with activation energies of 0.12–0.15 eV [9–16]. Replacing hydrogen with deuterium during interface passivation reduces NBTI [10]. However, effect of such replacement on NBTI is far less dramatic than that in hot carrier degradation.
- (f) Both interfaces as well as gate oxide processing influences NBTI robustness. The presence of water [26], boron [9,11], and nitrogen [17] near the Si/SiO₂ interface can modify NBTI substantially.
- (g) Although NBTI is usually associated with interface traps only, there are some reports that indicate the number of positive charges in the oxide bulk equal the number of interface traps under NBTI stress conditions [1,5,26] indicating correlated generation. Whether this equality is preserved during the entirety of the degradation is debatable.

None of the existing NBTI models are comprehensive enough to consistently explain the above observations within a common framework. The goal of this paper is to develop such a comprehensive framework by generalizing the classical reaction–diffusion (R–D) model [5,13–16] for N_{IT} generation. This would allow us to identify the roles of hole density, hydrogen diffusion, oxide thickness, and interface quality that affect NBTI reliability. We will use the above experimental observations to systematically test the R–D model whenever possible and augment the model whenever necessary.

2. The standard R–D model of NBTI

2.1. Description of the R–D model

Till date, the R–D model is the only model that can interpret the power-law dependence of interface trap generation during NBTI, as discussed in Section 1.2(a), without making any a priori assumption regarding the distribution of interface bond-strengths. As discussed in Ref. [5,13,14], the model assumes that when a gate voltage is applied, it initiates a field-dependent reaction at the Si/SiO₂ interface that generates interface traps by breaking the passivated Si–H bonds. Although the exact mechanism that causes such bond dissociation remains unspecified in the original model [5] and a number of groups are exploring the mechanics of trap generation by first-principle calculations [29–31], some have suggested [13–17] (and we will discuss this in detail in Section 3) that this dissociation is preceded by the capture (via field assisted tunneling) of inversion layer holes by the Si–H covalent bonds. This weakens the existing Si–H covalent bond [24], which is easily broken at higher

temperatures. The newly released hydrogen diffuses away from the interface, leaving behind a positively charged interface state (Si^+) that is responsible for higher threshold voltage and lower transconductance. This process is schematically explained in Fig. 1. The R–D model is described by the following equations:

$$dN_{\text{IT}}/dt = k_{\text{F}}(N_0 - N_{\text{IT}}) - k_{\text{R}}N_{\text{H}}N_{\text{IT}} \quad (x = 0) \quad (1a)$$

$$dN_{\text{IT}}/dt = D_{\text{H}}(dN_{\text{H}}/dx) + (\delta/2)dN_{\text{H}}/dt \quad (0 < x < \delta) \quad (1b)$$

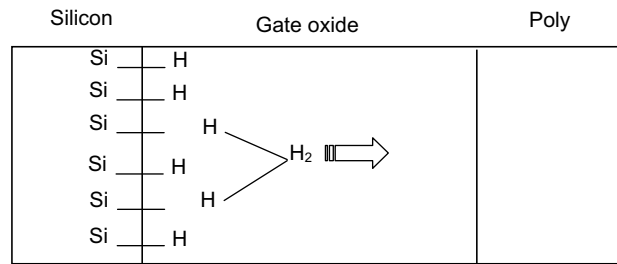
$$D_{\text{H}}(d^2N_{\text{H}}/dx^2) = dN_{\text{H}}/dt \quad (\delta < x < T_{\text{PHY}}) \quad (1c)$$

$$D_{\text{H}}(dN_{\text{H}}/dx) = k_{\text{P}}N_{\text{H}} \quad (x > T_{\text{PHY}}) \quad (1d)$$

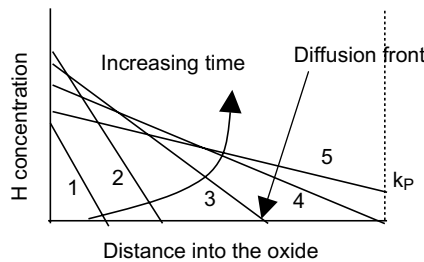
where $x = 0$ denotes the Si– SiO_2 interface and $x > 0$ is (in the oxide) towards the gate, N_{IT} is the number of interface traps at any given instant, N_0 is the initial number of unbroken Si–H bonds, N_{H} is the hydrogen concentration, k_{F} is the oxide field dependent forward dissociation rate constant, k_{R} is the annealing rate constant, D_{H} is the hydrogen diffusion coefficient, δ is the interface thickness, T_{PHY} is the oxide thickness and k_{P} is

the surface recombination velocity at the oxide/poly-silicon interface. Note that no field dependent term in Eq. (1c) means that the diffusing species is assumed neutral. Also, the microscopic details of the trap generation and trap annealing processes that occur within a few angstrom of the Si/ SiO_2 interface (i.e., δ) are assumed hidden in the constants k_{F} and k_{R} . For now, these parameters are measured and parameterized in terms of device variables like stress bias, stress temperature, and oxide thickness (see Section 3). As Si–H bond-dissociation process is better understood [29–31], one may be able to predict k_{F} and k_{R} from first principle.

Before we proceed to compare the solutions of the above R–D model to experimental data, note that this model attributes V_{T} shift during NBTI entirely to N_{IT} generation and contributions from slow traps, if any, is not separately distinguished. Moreover, note that measured ΔV_{T} , especially for thick oxides at high voltages, contains contributions from both interface traps due to NBTI and bulk traps (N_{OT}) due to hot-holes, e.g., $\Delta V_{\text{T}} = q/C_{\text{OX}}(N_{\text{IT}} + N_{\text{OT}})$ (q is electronic charge and C_{OX} is oxide capacitance) [15]. Therefore, bulk-trap contributions must be separated before R–D model predictions can be compared with measurements. Two



(a)



(b)

Fig. 1. (a) Schematic description of the reaction–diffusion model used to interpret interface-trap generation during NBTI stress. Broken Si–H bonds at the Si– SiO_2 interface create Si^+ (interface traps) and H. Initial interface-trap generation rate depends on Si–H bond dissociation (reaction), while the later rate depends on H removal (diffusion). Some H converts to H_2 and since H is a slow diffuser, H removal is triggered by H_2 diffusion (see Section 3). (b) Hydrogen profile in the oxide during NBTI stress. The area under the hydrogen profile equals generated interface traps (see Section 2). The shape and sequence of these profiles reflect the reaction-limited (1,2) and diffusion-limited (3,4) regimes. Once the diffusion front reaches the SiO_2 /poly-interface (5), faster diffusivity of H_2 in poly ensures that the concentration of H_2 at the poly-oxide interface will change only modestly. This would lead to a sharper gradient in the hydrogen profile resulting in faster H removal from the Si/oxide interface, and would trigger enhanced interface trap generation.

separate approaches can be taken to de-embed the effects of bulk traps from measured threshold voltage. First, one can theoretically estimate the bulk-trap contribution for a given oxide thickness for every stress voltage and temperature [15,28] and subtract this contribution to extract the N_{IT} contribution from measured ΔV_T data. Second, during measurement one can carefully avoid such “non-NBTI” contributions to ΔV_T due to bulk traps by carefully optimizing the accelerated stress conditions and demonstrating a good correlation between N_{IT} generation and V_T shift [16]. We will use both approaches in appropriate in the following discussion.

2.2. Solution of the R–D model

2.2.1. Stress phase

The general solution of the R–D model in the stress phase anticipates five different regimes of time-dependent interface trap generation [14], as shown in Fig. 2.

During the very early phase, both N_{IT} and N_H are small with negligible hydrogen diffusion, and therefore by Eq. (1a), $N_{IT} \sim k_F N_0 t$. During the second phase, the forward and reverse reactions in Eq. (1a) are both large but approximately equal (i.e., $k_F N_0 \sim k_R N_H(x=0) N_{IT}$) and also all the hydrogen released are still at the interface so that $N_H(x=0) = N_{IT}$. Therefore, Eq. (1a) suggests $N_{IT} \sim (k_F N_0 / k_R)^{1/2} t^0$. For a particular set of parameters relevant for NBTI in modern MOSFETs, both these phases last up to a few microseconds, and therefore are not observed in typical NBTI measurements.

The phase of NBTI that is most readily observed in typical measurements is when hydrogen diffusion (Eqs. (1b) and (1c)) begins to control the trap generation process. In this regime, Eq. (1c) suggests $x = (D_H t)^{1/2}$ and by Eq. (1b), $dN_{IT}/dt \sim D_H (dN_H/dx) \sim D_H N_H(x=0) /$

$(D_H t)^{1/2}$, so that $N_H(x=0) \sim (t/D_H)^{1/2} (dN_{IT}/dt)$. Substituting this in Eq. (1a) and assuming that net trap generation is so slow that it is negligible *in comparison* to the large fluxes on the right hand side of Eq. (1a) (e.g., $dN_{IT}/dt \sim 0$), we find

$$N_{IT} \sim (k_F N_0 / k_R)^{1/2} (D_H t)^{1/4} \quad (2a)$$

The above result is important enough to re-derive it in a slightly different, but perhaps more instructive manner. The total number of interface traps ever generated must equal to the number of hydrogen released, therefore $N_{IT} = 1/2 N_H(x=0) (D_H t)^{1/2}$ (by assuming linear hydrogen profile) and by Eq. (1a) $N_H(x=0) N_{IT} \sim k_F N_0 / k_R$. Using these two relations, we can re-derive Eq. (2a), and also find that in this regime, while the total number of released hydrogen is increasing at the same rate as N_{IT} , the density of hydrogen at the interface is decreasing as $N_H(x=0) \sim t^{-1/4}$ (see profiles 2,3,4 in Fig. 1(b)).

The NBTI time dependence of $(t^{1/4})$ derived above is based on analytical simplifications—numerical solution of Eqs. (1a)–(1c) shows that the exponent can be somewhat larger ($n \sim 0.25$ – 0.28). Also, $x = (D_H t)^{1/2}$ from Eq. (1c) implies that the hydrogen diffusion front is yet to reach the poly-silicon interface. Therefore, this is the only regime that need be considered for very thick oxides [5]. Finally, had the diffusing species been positively charged (proton?), the time dependence would have been $t^{1/2}$ [13]—which is not observed during this phase of NBTI degradation.

Once the hydrogen reaches the oxide/poly-interface (profile 5, Fig. 1(b)), according to Eq. (1d) the incoming flux $D_H (N_H(x=0) - N_H(x = T_{PHY})) / T_{PHY} = k_P N_H(x = T_{PHY})$, so that $N_H(x=0) = (1/k_P + T_{PHY}/D_H) (dN_{IT}/dt)$ by Eq. (1b). Substituting this in Eq. (1a) and assuming the trap generation rate is slow compared to the generation and annealing fluxes in Eq. (1a), (i.e., $dN_{IT}/dt \sim 0$), we find [14,16]

$$N_{IT} = A (k_F N_0 / k_R)^{1/2} (D_H t)^{1/2} \quad (2b)$$

with $A = [2(D_H/k_P + T_{PHY})]^{-1/2}$. This regime ($n = 1/2$) is likely to be observed in thinner oxides where the diffusion front would reach the poly-interface within measurement window. This is explicitly shown in Fig. 3 [16], where measured V_T shift (for thinner oxides at high stress temperature) breaks off from the initial trend and increases at longer times. However, although the break is significant and clear, note that the value of the post-break exponent, n , is not equal to $1/2$, as one might anticipate from Eq. (2b). We *speculate* that if the diffusion channels in poly-silicon grain boundaries saturate over time, the surface recombination velocity, k_P , will reduce over time as well. This will cause the N_H concentration to build-up at the poly/oxide interface (Fig. 1, profile 5), with a corresponding reduction of hydrogen diffusion away from the Si/oxide interface. This would in

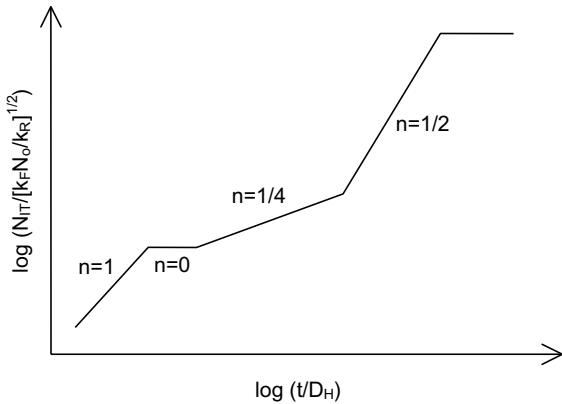


Fig. 2. Five regimes of time dependent interface-trap generation as obtained from the general solution of the reaction–diffusion equations during NBTI stress.

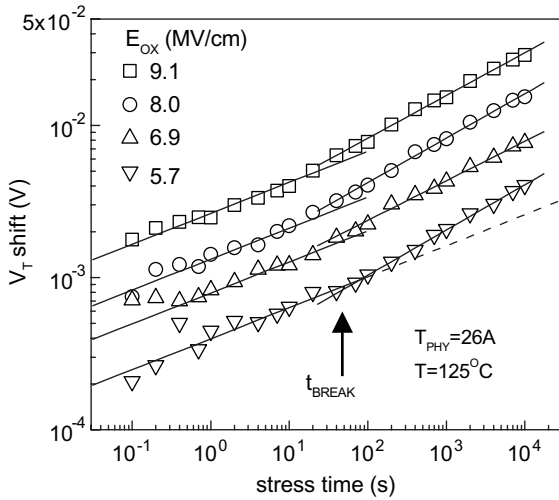


Fig. 3. Time evolution of threshold voltage shift of a 26 A PMOS during NBTI stress at 125 °C [16]. The degradation rate is small initially but breaks off and increases at longer time. The break time and post-break slope are insensitive to oxide field (gate voltage), suggesting the long-time enhancement is not a bulk-trap activated process [15]. Such enhancement is anticipated by the R–D model (Section 2.2, Figs. 1 and 2). Absence of field acceleration of the breakpoint suggests neutral diffusion species.

turn reduce N_{IT} below the values predicted by Eq. (2b). Therefore, although the break (in degradation rate) will occur once the hydrogen front reaches the poly/oxide interface, the exponent n could be smaller than 1/2.

Finally, when all the Si–H bonds are broken, $N_{IT} \sim N_0 = \text{Const}$. It is unlikely that we can observe this phase of NBTI degradation experimentally, because other failure modes like TDDDB are likely to cause oxide breakdown in the meantime.

2.2.2. Annealing phase

Apart from the five-phase degradation discussed above, another key prediction of the R–D model is similar multi-phase annealing of the interface traps (created during the stress phase) once the stress is removed (see Section 1.2(b)). Once the stress is removed, the dissociation of the Si–H bond that forced forward diffusion of hydrogen away from the interface no longer exists, therefore the hydrogen can now diffuse back and recombine with silicon dangling bonds restoring them to their passive Si–H state.

This diffusion controlled annealing phase (analogous to the diffusion-controlled generation phase described by Eq. (2a)) can be treated analytically in the following manner [20]. Assume that at the end of the stress phase at time t_0 , the number of traps generated is $N_{IT}^{(0)}$ and the density of interface hydrogen is $N_H^{(0)}$, so that $N_{IT}^{(0)} = 1/2N_H^{(0)}(D_H t_0)^{1/2}$ (assuming a linear hydrogen profile).

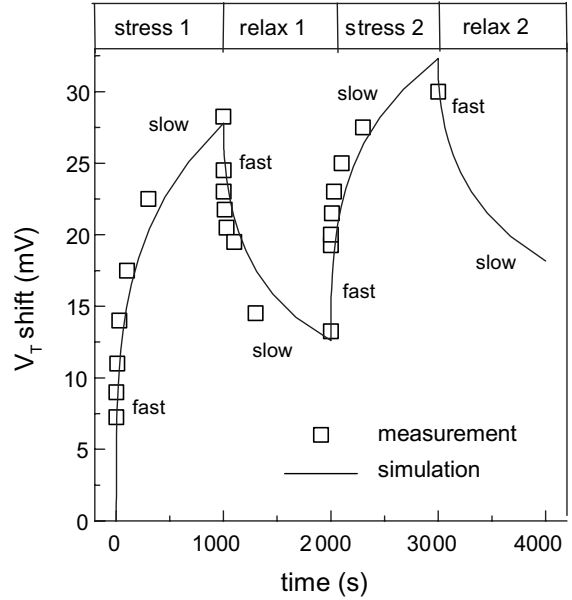


Fig. 4. The R–D model is solved numerically for two cycles of interrupted stress (stress 1000 s, relaxation 1000 s). The simulation results (continuous line) agree well with measured data (symbols) [21]. The parameters used are $k_F = 10^{-2}/s$, $k_R = 10^{-18} \text{ cm}^3/s$, $N_0 = 1.24 \times 10^{14}/\text{cm}^2$ and $T_{PHY} = 1.3 \text{ nm}$. Each 1000 s segment is characterized by a fast phase (first few seconds) followed by a slower phase of interface-trap generation, which relate to reaction-limited and diffusion-limited regimes, respectively.

During the recovery phase $N_{IT}^{(*)}$ traps are annealed at time $(t + t_0)$ (the start of the annealing phase defines the time-origin here), so that $N_{IT}^{(*)} = 1/2N_H^{(*)}(\xi D_H t)^{1/2}$ ($\xi = 1/2$ for one sided diffusion). With $k_F = 0$, Eq. (1a) can be rewritten as $k_R [N_H^{(0)} N_H^{(*)}] [N_{IT}^{(0)} - N_{IT}^{(*)}] \sim 0$. Assuming that the original hydrogen front continues to diffuse as $[D_H(t + t_0)]^{1/2}$, one finds

$$N_{IT} = N_{IT}^{(0)} [1 - (\xi t/t_0)^{1/2} / (1 + t/t_0)^{1/2}] \quad (t > 0) \quad (2c)$$

Such annealing effects have indeed been reported in literature by a number of groups [21–23,25]. Fig. 4 shows that Eqs. (2a) and (2c) can interpret the stress and the relaxation phases of the NBTI degradation well. This is yet another indication that the diffusing species is charge neutral, because a positively charged species would diffuse asymmetrically with stress (negative) bias and anneal (zero) bias conditions, which would have made Eq. (2c) inconsistent with experimental data.

2.3. Discussion of the R–D model

The R–D model successfully explains both the power-law dependence of interface trap generation and the

mechanics of interface trap annealing—the two key features of NBTI as discussed in Section 1.2(a) and (b). It does so without requiring that holes be present for NBTI degradation (although it does not forbid it either) and apart from requiring that the diffusing species be neutral, the theory places no restriction on the nature or type of diffusing species involved. Furthermore, it offers no quantitative predictions regarding the field or temperature dependence of the NBTI phenomena. Therefore, one cannot optimize operating or processing conditions to reduce NBTI effects using the basic R–D model. In Sections 3.1 and 3.2 below we will generalize of the R–D model to predict the field and temperature dependence of the NBTI phenomena. Moreover, the predictions of precise power-exponents (e.g., $n = 1/4$) also make the R–D model difficult to reconcile with wide variety of exponents observed in various experiments. Failure to account for wide variety of power-exponents is another reason to generalize the standard model. We will take this up in Section 3.2.4.

3. Enhancement of the R–D model

3.1. Model for NBTI field (or voltage) dependence

According to the R–D model, the field dependence of the NBTI phenomena must be confined to the processes at the interface (k_F and k_R) because the diffusing species is charge neutral. We anticipate that the forward dissociation constant, k_F , should depend on the number of holes (p), their ability to tunnel to the Si–H bonds (T_{COEFF}), the capture cross-section of the Si–H bonds (σ_0), and any field dependence of Si–H bond dissociation (B), so that $k_F(\dots, E_{\text{OX}}) = B\sigma_0 p T_{\text{COEFF}}$. We discuss these individual processes below.

3.1.1. Role of holes and field-dependence through hole density

The presence of holes have been linked to NBTI degradation from the beginning of the study of this process—in part because the effect was first observed in PMOS devices in inversion. However, the lack of degradation in NMOS devices stressed in accumulation with voltages similar to those used for PMOS NBTI tests have sometimes been used to argue against the role of holes in NBTI degradation [12]. However, since we must compare the NBTI degradation at comparable hole densities, the surface field must be same—which means that the NMOS must be stressed approximately one volt higher compared to PMOS devices for comparable NBTI effects (to take care of flatband voltage difference). Fig. 5 shows that this is indeed the case—this explains the experimental observation discussed in Section 1.2(c) and leads us to conclude that the presence of holes play a key role in NBTI degradation consistent

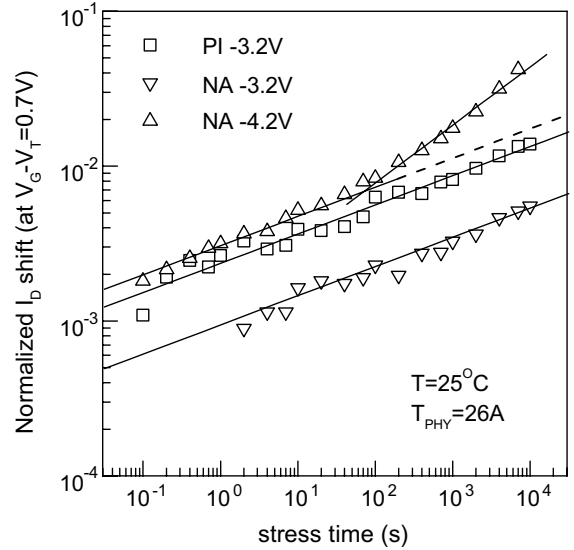


Fig. 5. Time evolution of (normalized) drain current shift for PMOS inversion and NMOS accumulation NBTI stress. Degradation at low V_G for all time and high V_G at shorter time is due to interface-trap generation, while long-time enhanced degradation at high V_G is due to bulk-trap generation [15,16]. Under identical oxide field ($V_G|_{\text{inv}} = V_G|_{\text{acc}} - 1$ V to account for difference in flatband voltage) PMOS inversion and NMOS accumulation (short time) stress show similar degradation, which characterizes the role of holes for NBTI.

with the observation in Section 1.2(c). The density dependence can be summarized as $p = [\text{COX}(V_G - V_T)] \sim E_{\text{OX}}$.

3.1.2. Role of hole capture and its field-dependence

Once the holes are available, they must tunnel through the interface layer of 1–2 Å to be captured by the Si–H bonds. The field dependent tunneling coefficient depends exponentially on the local electric field at the interface (e.g., $T_{\text{COEFF}} \sim \exp(E_{\text{OX}}/E_0)$, although other forms are also suitable [1,2,15,23]). Fig. 6 shows that this is indeed the case for oxides of various thicknesses, confirming the general validity of the approach [24–26,28].

3.1.3. Role of field-dependent bond-dissociation

A key unknown is the influence of the electric field on the dissociation of the Si–H bond itself. Assuming that such modification will be associated with a reduction in the barrier height of the bond dissociation, we anticipate that there may be field dependent activation energy associated with this process. Such field dependence of bond dissociation has not been experimentally demonstrated yet, but may have to be included in future models.

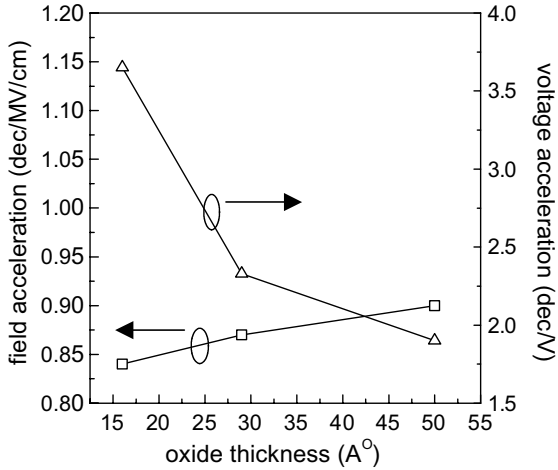


Fig. 6. Voltage- and field-acceleration factors (γ_V and γ_E , respectively) plotted as a function of oxide thickness. The thickness independence of the field acceleration factor suggests lifetime of the form $\tau \sim \exp(\gamma_E E_{OX})$. Since $N_{IT} \sim \tau^n$, therefore $N_{IT} \sim \exp(E_{OX}/E_0)$ with $E_0 = 1/n\gamma_E$ is the form that describes the NBTI data well.

Taken together, the field dependence is given by

$$k_F(\dots, E_{OX}) = B\sigma_0 E_{OX} \exp(E_{OX}/E_0). \quad (3a)$$

3.1.4. Role of bulk traps

Finally, as discussed earlier in the paper, it is important to realize that at high negative electric fields (especially for thicker gate oxides that involve large stress voltages) NBTI data are often contaminated by generation of bulk traps [15]. Such contribution from bulk traps must be isolated before a universal representation of k_F (Eq. (3a)) can be found. According to the Anode Hole Injection model, at large negative voltages in both PMOS and NMOS transistors, electrons are injected from the poly-gate into the silicon substrate. These energetic electrons in turn produce hot holes that are injected back into the oxide-creating bulk defects within the oxides [28]. As shown in Fig. 7, this contribution from bulk defect densities can easily be identified and isolated by its time dependence ($n = 1/2$) and its unique voltage and temperature dependence (for details, see [15]). Once such corrections are accounted for, we find that field dependence of Eq. (3a) defines the forward reaction rate well and is consistent with the observation in Section 1.2.

3.2. Model for NBTI temperature dependence

3.2.1. Concept of universal scaling

Eq. (2a) provides a universal scaling relationship that can be used to decouple the field- and temperature-dependence of NBTI. This equation predicts that if the stress time is measured in units of $(1/D_H)$ and if the trap density is normalized by $(k_F N_0/k_R)^{1/2}$, then the resulting curve should be universal. Moreover, since D_H is a function of temperature alone and is independent of electric field, and since $(k_F N_0/k_R)^{1/2}$ is field dependent, but temperature independent (approximately, see below), the NBTI curves at different temperatures (but at a constant electric field) can be scaled along the time axis to identify the activation energy of the diffusion process [14–16]. Fig. 8 highlights the scaling methodology and shows that universal scaling holds for NBTI data measured under a wide range of stress bias and temperature.

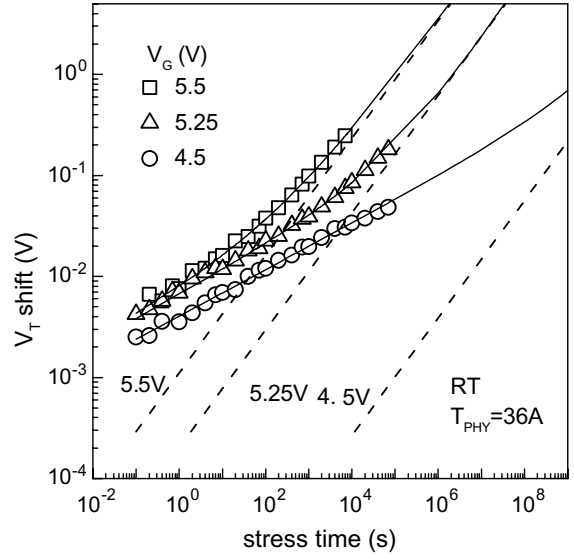


Fig. 7. Experimental and theoretical time evolution of threshold voltage shift data for a wide range of stress V_G . The calculated bulk trap contributions are shown by dashed lines [15]. Theory matches well with experiment. Bulk trap generation at large oxide fields (high V_G) affects overall threshold voltage shift and must be taken into account for properly estimating the forward reaction rate of the R–D model.

Although the hole density (p) and the hole capture cross-section (σ) are largely temperature independent, the dissociation and annealing of Si–H bonds (k_F and k_R) and the diffusion of hydrogen (D_H) through the oxide are not. Therefore, based on Eq. (2a), the following form must describe N_{IT} :

3.2.2. Activation energy of the diffusion process

with net activation energy of

$$N_{IT} = [k_{F0}(\dots, E_{OX})N_0/k_{R0}]^{1/2} \exp[(-E_a(k_F) + E_a(k_R)) / 2kT] [D_{H0} \exp(-E_a(D_H)/kT)t]^{1/4}, \quad (3b)$$

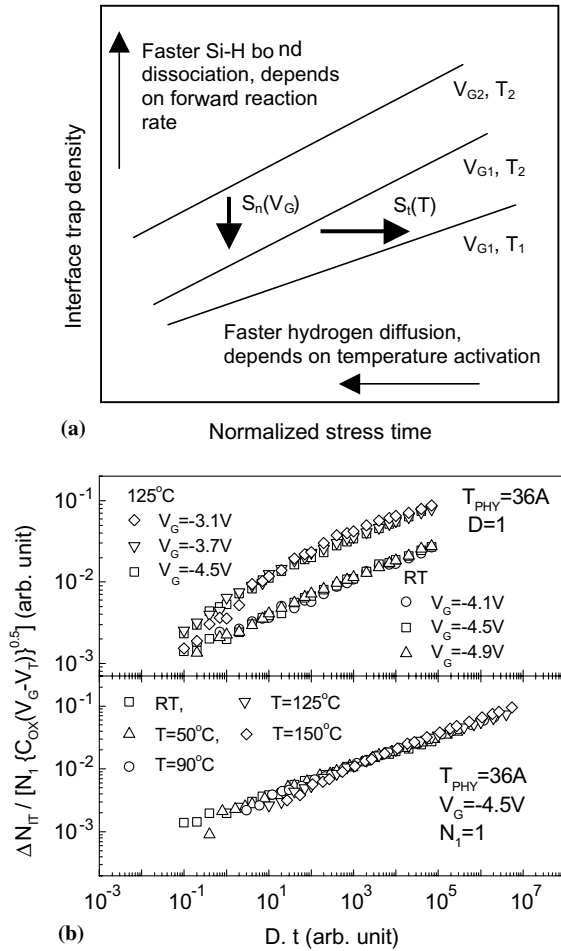


Fig. 8. (a) Universal scaling scheme for generated interface traps versus time data. Density axis reflects higher Si–H bond dissociation (reaction) while time axis denotes faster hydrogen diffusion. Assuming neutral diffusion species, the reaction and diffusion components can be independently scaled by voltage dependent (time independent) and time dependent (voltage independent) functions respectively. (b) Universality of scaled generated interface traps versus time data. Data taken under a wide bias and temperature range validates the scaling hypothesis.

We use two methods to determine the activation energies. The first method uses the time-scaling idea discussed above that allows us to determine E_a directly (Fig. 8). The second technique uses the fact that the change in power-law exponent from $n \sim 1/4$ to $n \sim 1/2$ (see Fig. 3) is determined by the time hydrogen front reaches the poly-interface, i.e., $D_H = T_{PHY}^2(1/t_{BREAK})$ (see Fig. 2). Plotting $(1/t_{BREAK})$ as a function of temperature, one determines the activation energy of the diffusion process, $E_a(D_H)$. Fig. 9 shows that activation energies determined by both methods are essentially identical, demonstrating that while the barrier heights for dissociation and annealing of Si–H bonds may

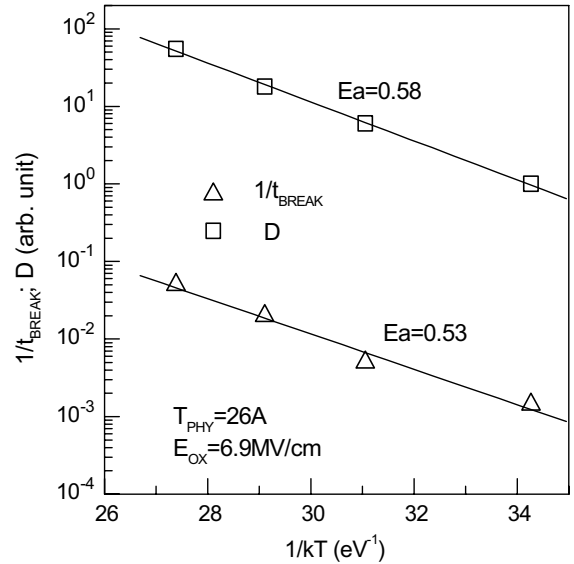


Fig. 9. Temperature activation of $1/t_{BREAK}$ and D_H . The D_H is calculated from the universal scaling of temperature dependent (fixed oxide field) generated interface trap data. Obtained activation energy supports neutral H_2 diffusion [32].

themselves be high, the difference $[-E_a(k_F) + E_a(k_R)]/2$ must be small. Moreover, the activation energy of 0.5–0.6 eV obtained by both these methods indicates neutral H_2 diffusion [32]. Therefore, the temperature activation of NBTI process is essentially determined by that of the diffusion process, i.e., $E_a(NBTI) \sim E_a(D_H)/4 = 0.12\text{--}0.15$ eV-consistent with Section 1.2(e).

3.2.3. Debate regarding the nature of the diffusing species

Although there is a debate regarding the exact diffusion species ($-OH$, H , and H_2O have been suggested), the activation energy of the diffusing species ($E_a(D_H) = 4E_a(NBTI) \sim 0.5$ eV) appears to indicate that the diffusing species is hydrogen, most probably a molecular species [32]. Moreover, the experiment in Ref. [10], when viewed in light of Eq. (2a), also supports the hydrogen hypothesis: When hydrogen annealing of Si-dangling bonds to create Si–H bonds is replaced with deuterium annealing to create Si–D bonds, Eq. (2a) anticipates that $N_{IT}(D)/N_{IT}(H) = (D_D/D_H)^{1/4} = (m_D/m_H)^{1/8}$ (m_H and m_D being the effective mass of hydrogen and deuterium respectively), making the dependence finite but weak [10]. Note that if $-OH$ were the diffusing species, the difference in masses of $-OH$ and $-OD$ would have been small, resulting in essentially indistinguishable impact of deuterium annealing. The lack of strong deuterium effect also highlights the chemical nature of the Si–H (or Si–D) bond breaking as opposed to the kinetic nature of Si–H bond breaking during HCl experiments. Such chemical

dissociation of Si–H bonds is consistent with the R–D model.

3.2.4. Hypothesis regarding the non-standard NBTI exponents

The R–D model predicts very precise power-law exponent that is independent of temperature, time, or oxide thickness (e.g., $n = 1/4$ analytically, or $n = 0.25$ – 0.28 numerically). However, as reported in the literature [7–26] and summarized in Section 1.2(a), the experimentally measured power exponent has a wide range that depends on oxide thickness, temperature, and time. One reason behind this wide range of observed power exponent is due to the NBTI recovery that takes place during the delay between stopping the stress for measurement and the start of next stress phase.¹ The other reason for such discrepancy lies in the assumption of constant time-independent diffusion coefficient for hydrogen diffusion in the amorphous oxide as discussed below.

Since the assumption of constant diffusion coefficient is only appropriate for isotropic media with spatially and temporally uniform hopping rates, this may not be appropriate for diffusion in amorphous media where the hopping distances and hopping times are exponentially distributed [33–35]. Therefore, although the diffusion process can be described by a constant coefficient at the earliest phases, the diffusive particles soon find themselves trapped in states with exponentially long release times, which in turn slows the overall diffusion process. Since such dispersive diffusion is time dependent (reflecting broad distribution of release time), $D_H = D_{H0}(\omega_0 t)^{-a}$ [36,37]. Therefore, using the same argument we used to derive Eq. (2a), we find that the interface trap density increases as $N_{IT} \sim t^{n-a/4}$. In other words, such dispersive transport will give rise to effective exponents n' that is less than n predicted by the standard R–D model (i.e., $n' = n - a/4$). Assuming that $a \sim 0.1$ – 0.2 [36,37], and initial $n = 0.25$, dispersive power-exponents of the order of $n' \sim 0.225$ – 0.20 are expected. Indeed such lower power law exponents have been reported in the literature [16].

¹ The absolute degradation exponents is likely to be somewhat lower than the measured exponents reported here ($n \sim 0.32$ – 0.23 in Fig. 10), because there is a finite time-delay at the end of each stress section before measurements are made. Since the ratio of the relaxation time (fixed in this case) to stress time (increasing logarithmically) determines the degree of relaxation (Eq. (2c)), the effective relaxation for each data-point decreases as stress progresses. This makes the measured exponents slightly higher compared to what an absolute no-delay measurement would indicate. Ideally, such no-delay, absolute exponents should be compared with the various theoretical predictions presented in this paper ($n \sim 0.27$ – 0.2).

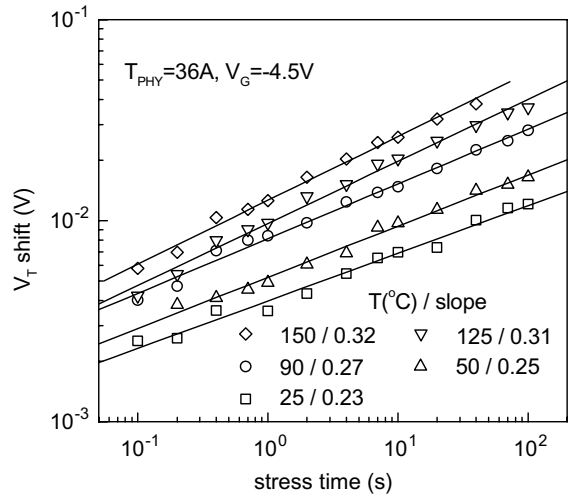


Fig. 10. Time evolution of threshold voltage shift under various stress temperature (fixed oxide field). Degradation exponents increases with increase in temperature.

One of the consequences of this dispersive diffusion in amorphous oxide is the temperature and time dependence of power-exponents observed in measured data. At lower temperatures, the diffusion is slow with lower hopping probability. Therefore, at a given time and a given site, diffusing hydrogen have time to relax to deeper states with long release times. Therefore, at lower-temperature, the observed power-exponents are typically lower, as shown in Fig. 10. At higher temperature, the hopping probability is higher, therefore initially the probability of getting trapped into deep level is reduced. This explains the higher value of degradation exponent at higher temperature (see Fig. 10, [38]). In time however a fraction of hydrogen will eventually find themselves trapped in deeper levels with long release time, making the asymptotic n' value similar to that of the low-temperature value [15].

The hypothesis discussed above, if supported by other experiments, would imply that more amorphous the oxide, the better is its NBTI performance, because deep level trapping with long release time would reduce NBTI power-exponent. This provides a technique to control NBTI degradation by processing changes for gate oxide deposition. Clearly more numerical and experimental work is necessary to support these ideas.

4. Conclusions

In this paper, we have discussed a theory of NBTI that uses the theoretical framework of R–D model to encapsulate the field, temperature, and processing dependencies of the NBTI phenomena. This overall

framework allows one to discuss and optimize the operating and processing conditions in a manner that is globally consistent and technologically relevant.

Despite its successes in explaining experimental observation (Section 1.2(a)–(e)), the model presented here still lacks systematic description of the role of boron and nitrogen at the interface. Such processing details will surely affect the magnitudes of both k_F and k_R (but most probably will not modify the functional dependence on field and temperature as discussed in this paper). One of the key goal of our future work would be to clarify the role of such processing changes on NBTI performance.

Acknowledgements

We would like to thank P. Bharath Kumar for technical contributions, T. Nigam, P. Mason, J. Vasi and R. Lal for many interesting discussions and S.J. Hillenius for encouragement and support.

References

- [1] Deal BE, Sklar M, Grove AS, Snow EH. Characteristics of the surface-state charge (Q_{ss}) of thermally oxidized silicon. *J Electrochem Soc* 1967;114:266.
- [2] Strain RJ, Goetzberger A, Lopez AD. On the formation of surface states during stress aging of thermal Si–SiO₂ interfaces. *J Electrochem Soc* 1973;120:90.
- [3] Frohman-Bentchkowsky D. A fully decoded 2048-bit electrically programmable FAMOS readonly memory. *IEEE J Solid State Circuits* 1971;6(5):301–6.
- [4] Nakagiri M. *Jpn J Appl Phys* 1974;13:1619.
- [5] Jeppson KO, Svensson CM. Negative bias stress of MOS devices at high electric fields and degradation of MOS devices. *J Appl Phys* 1977;48:2004–14.
- [6] Hu C, editor. *Nonvolatile semiconductor memories: technology, design, and applications*. Piscataway, NJ: IEEE Press; 1991.
- [7] Chaparala P, Shibley J, Lim P. Threshold voltage drift in p-MOSFETs due to NBTI and HCI. In: *Proc Int Reliability Workshop*, 2000. p. 95–7.
- [8] La Rosa G, et al. NBTI—channel hot carrier effects in p-MOSFETs in advanced CMOS technologies. In: *Proc Int Reliability Phys Symp*, 1997. p. 282–6.
- [9] Uwasawa K, Yamamoto T, Mogami T. A new degradation mode of scaled p+ polysilicon gate p-MOSFETs induced by bias temperature instability. In: *Proc Int Electron Device Meet*, 1995. p. 871–4.
- [10] Kimizuka N, Yamamoto T, Mogami T, Yamaguchi K, Imai K, Horiuchi T. The impact of bias temperature instability for direct tunneling ultra-thin gate oxide on MOSFET scaling. In: *Proc VLSI Tech Symp*, 1999. p. 73–4.
- [11] Yamamoto T, Uwasawa K, Mogami T. Bias temperature instability in scaled p+ polysilicon gate p-MOSFETs. *IEEE Trans Electron Devices* 1999;46(5):921–6.
- [12] Makabe M, Kubota T, Kitano T. Bias temperature degradation of p-MOSFETs: mechanism and suppression. In: *Proc Int Reliability Phys Symp*, 2000. p. 205–9.
- [13] Ogawa S, Shiono N. Generalized diffusion–reaction model for the low-field charge build up instability at the Si–SiO₂ interface. *Phys Rev B* 1995;51(7):4218–30.
- [14] Alam M, Weir B, Silverman P. The prospect of using thin oxides for silicon nano transistor. In: *Proc Int Workshop on Gate Insulator*, 2001. p. 10–3.
- [15] Mahapatra S, Alam MA. A predictive reliability model for PMOS bias temperature degradation. In: *Proc Int Electron Device Meet*, 2002. p. 505–9.
- [16] Mahapatra S, Bharat Kumar P, Alam MA. A new observation of enhanced bias temperature instability in thin gate oxide p-MOSFETs. In: *Proc Int Electron Device Meet*, 2003. p. 337–41.
- [17] Mitani Y, Nagamine M, Satake H, Toriumi A. NBTI mechanism in ultra-thin gate dielectric-nitrogen-originated mechanism in SiON. In: *Proc Int Electron Device Meet*, 2002. p. 509–12.
- [18] Schroder DK, Babcock JA. Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing. *J Appl Phys* 2003;94:1–18.
- [19] Nishida Y, et al. SoC CMOS technology for NBTI/HCI immune I/O and analog circuits implementing surface and buried channel structures. In: *Proc Int Electron Device Meet*, 2001. p. 869–72.
- [20] Alam M. A critical examination of the mechanics of dynamic NBTI for p-MOSFETs. In: *Proc Int Electron Device Meet*, 2003. p. 346–9.
- [21] Chen G, et al. Dynamic NBTI of PMOS transistors and its impact on device lifetime. In: *Proc Int Reliability Phys Symp*, 2003. p. 196–202.
- [22] Huard V, Monsieur F, Ribes G, Bruyere S. Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs. In: *Proc Int Reliability Phys Symp*, 2003. p. 178–82.
- [23] Tsujikawa S, et al. Negative bias temperature instability of pMOSFETs with ultra-thin SiON gate dielectrics. In: *Proc Int Reliability Phys Symp*, 2003. p. 183–8.
- [24] McPherson JW, Khamankar RB, Shanware A. Complementary model for intrinsic time dependent dielectric breakdown in SiO₂ dielectrics. *J Appl Phys* 2000;88(9):5351–9.
- [25] G. La Rosa, IRPS Tutorial, 2003.
- [26] Blat CE, Nicollian EH, Poindexter EH. Mechanism of negative bias temperature instability. *J Appl Phys* 1991; 69:1712.
- [27] Liu C-H et al. Mechanism of threshold voltage shift (ΔV_{th}) caused by negative bias temperature instability (NBTI) in deep submicron pMOSFETs. *Jpn J Appl Phys* 2002;41: 2423–5.
- [28] Alam M, Bude J, Ghetti A. Field acceleration for oxide breakdown—can an accurate anode hole injection model resolve the E vs. $1/E$ controversy? In: *Proc Int Reliability Phys Symp*, 2000. p. 21–6.
- [29] Soon J et al. Study of negative-bias temperature-instability-induced defects using first-principle approach. *Appl Phys Lett* 2003;83:3063–5.
- [30] Tan SS et al. Nitrogen-enhanced negative bias temperature instability: An insight by experiment and first-principle calculations. *Appl Phys Lett* 2003;82:1881–3.

- [31] Ushio J, Maruizumi T, Abdelghafar KK. Interface structures generated by negative-bias temperature instability in Si/SiO₂ and Si/SiO_xN_y interfaces. *Appl Phys Lett* 2002;81: 1818–20.
- [32] Reed ML, Plummer JD. Chemistry of Si–SiO₂ interface trap annealing. *J Appl Phys* 1988;63:5776–93.
- [33] Scher H, Montroll EW. Anomalous transit-time dispersion in amorphous solids. *Phys Rev B* 1975;12:2455–77.
- [34] Monroe D. Hopping in band tails far from equilibrium. In: Pollak M, Shklovskii B, editors. *Hopping transport in solids*. Elsevier Science Publishers B.V; 1991.
- [35] Zallen R. *The physics of amorphous solids*. John Wiley & Sons, Inc.; 1983.
- [36] Kakalios J, Street RA, Jackson WB. Stretched-exponential relaxation arising from dispersive diffusion of hydrogen in amorphous silicon. *Phys Rev Lett* 1987;59:1037–40.
- [37] Shkrob IA, Trifunac AD. Time-resolved EPR of spin-polarized mobile H atoms in amorphous silica: The involvement of small polarons. *Phys Rev B* 1996;54:15073–8.
- [38] Reddy V, et al. Impact of negative bias temperature instability on digital circuit reliability. In: *Proc Int Reliability Phys Symp*, 2002. p. 248–54.