Critical analysis of short-term negative bias temperature instability measurements: Explaining the effect of time-zero delay for on-the-fly measurements

Ahmad Ehteshamul Islam,[a] Haldun Kufluoglu, Dhanoop Varghese, and Muhammad Ashraful Alam
School of ECE, Purdue University, West Lafayette, Indiana 47907-1285

(Received 7 November 2006; accepted 19 January 2007; published online 21 February 2007)

Recently several groups have used the reaction-diffusion (R-D) model with H_2 diffusion in interpreting negative bias temperature Instability (NBTI) degradation. While the classical "H_2 R-D" model can interpret long-term NBTI behavior, it is inconsistent with short-term stress data obtained by recently developed ultrafast measurements and widely used on-the-fly measurements. Moreover, experimental data from various techniques are not consistent with each other. Here, the authors show that the H_2 R-D model must be generalized to consistently interpret NBTI *at all time scales*. The generalized model highlights the previously unappreciated role of time-zero delay in reconciling differences among the so-called delay-free on-the-fly measurements. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2695998]

Negative bias temperature instability (NBTI), a major degradation mechanism for *p*-type metal-oxide-semiconductor devices biased in inversion, results in variation of transistor parameters and is generally monitored through threshold voltage shift (ΔV_T). The origin of such degradation has been debated for the last few years, and it is now generally accepted that, for devices with SiO_2, plasma SiON, and thin thermal SiON gate dielectrics, NBTI degradation results mainly from depassivation of Si–H bonds at the Si/dielectric interface and resultant diffusion of hydrogen species into gate dielectric and poly-Si.[1–8] As such, several groups have used the reaction-diffusion (R-D) model to interpret NBTI degradation.[1–7] At long stress time ($t_{\text{stress}} > 10-100$ s), ΔV_T shows a power law behavior when plotted with respect to time ($\Delta V_T \sim At^n$) with a consistent time exponent (n) of $\sim 1/6$. This time exponent is independent of the measurement techniques used to monitor ΔV_T [e.g., ultrafast measurement (UFM),[9] "delay-free" on-the-fly I_{dlin} measurement (OTFM),[2,3,5,6,8] etc.]. The R-D model attributes the robust (constant over several decades in time[4]) long-term $n \sim 1/6$ exponent to the diffusion of molecular hydrogen (H_2).[3,7] This "H_2 R-D" model also provides a consistent interpretation of temperature and field dependencies of NBTI for long-term stress, as extensively studied for devices with SiO_2, plasma SiON, and thin thermal SiON dielectrics.[2,5,6,8]

Although the classical "H_2 R-D" model provides an excellent interpretation for long-term stress data, our analysis shows that (1) the predictions of this model is *inconsistent* with the short term, sub- 10 s, NBTI degradation [see Fig. 1(b)] and (2) in contrast to long-term degradation, the short-term time exponent of NBTI degradation depends on the measurement techniques. In this letter, we show that the inconsistency between theory ("H_2 R-D" model) and experiment arises from the implicit assumption in the "H_2 R-D" model that atomic to molecular hydrogen (H to H_2) transformation is instantaneous. Once this assumption is relaxed and the model is generalized to explicitly account for both H and H_2 ("H–H_2 R-D" model), one can consistently interpret NBTI data at all stress times (Fig. 1). We also trace the differences in short-term time exponents obtained using various techniques to the previously unappreciated role of time-zero delay for OTFM. Once the time-zero delay is accounted for, the "H–H_2 R-D" model explains experimental data at all time scales independent of measurement techniques (Fig. 2).

As stated earlier, *long-term* NBTI stress can be explained with the "H_2 R-D" model as shown in Fig. 1(b). This figure also signifies that H_2 diffusion (alone) cannot properly explain the short-term stress data with the parameters normally used in the model.[2,3,5–7] Whereas H_2 diffusive model

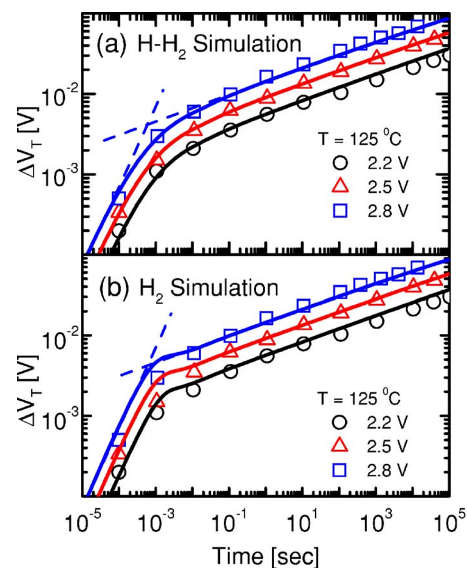

FIG. 1. (Color online) (a) Generalized R-D model (H–H_2 system) can explain the experimental trends (Ref. 9) at any stress time. R-D model parameters [consistent with literature (Refs. 3, 6, and 19): $k_f = 6 \times 10^{-3} E_c \exp(0.65 E_{\text{ox}})$ s^{-1} [E_c: oxide electric field due to mobile carriers; E_{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}, $D_{H_2} = 1.8126 \times 10^{-14}$ cm^2 s^{-1}, $k_H = 1.4 \times 10^{-3}$ cm^3 s^{-1}, $k_{H_2} = 95.4$ s^{-1}. (b) Classical H_2 R-D model (with large k_H and k_{H_2}) cannot explain short t_{stress} data consistently. The dashed asymptotes in (a) and (b) indicate that the transition from reaction to H_2 diffusion for H–H_2 system is slower compared to a classical H_2 system.

[a]Electronic mail: aeislam@purdue.edu

 © 2007 American Institute of Physics
Downloaded 25 May 2007 to 128.46.94.118. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

predicts a change in n from 1 (for reaction-limited region[1,7]) to 1/6 (for H_2 diffusion-limited region[3,7]) within approximately one order of time scale (see the dashed asymptotes in Fig. 1(b)), experiments indicate a wider transition region (Fig. 1), approximately three to four orders of magnitude.

The inability of the "H_2 R-D" model to explain the short-term stress condition indicates that the assumption of instantaneous formation of H_2 immediately after the dissociation of Si–H bond (creating interface traps), while appropriate for analysis of long-term degradation, may not be appropriate to interpret short-term data. This requires explicit consideration of H within the R-D framework, as this should be the first by-product after interface trap generation, before getting transformed to H_2.[3,7] Although neutral charge state may not be a stable form of atomic hydrogen, its transient formation is indeed possible.[10] Therefore, in the generalized R-D model, diffusion of both H and H_2 and $H \leftrightarrow H_2$ conversion are explicitly incorporated in the R-D framework by the following equations:[11]

$$\frac{dN_{IT}}{dt} = k_f(N_0 - N_{IT}) - k_r N_{IT} N_H^{(0)}, \quad (1)$$

$$\frac{dN_H}{dt} = D_H \frac{d^2 N_H}{dx^2} - k_H N_H^2 + k_{H_2} N_{H_2}, \quad (2)$$

$$\frac{dN_{H_2}}{dt} = D_{H_2} \frac{d^2 N_{H_2}}{dx^2} + \frac{1}{2} k_H N_H^2 - \frac{1}{2} k_{H_2} N_{H_2}. \quad (3)$$

In Eq. (1), k_f, k_r, N_0, N_{IT}, $N_H^{(0)}$ are defined as[1–3,11] the Si–H bond-breaking rate, Si–H bond-annealing rate, initial bond density available before stress, interface trap density, and hydrogen density at the Si/dielectric interface, respectively. In Eqs. (2) and (3), k_H and k_{H_2} represent the generation and dissociation rates of H_2; D_H and D_{H_2} represent the diffusion coefficients for H and H_2; N_H and N_{H_2} represent the concentrations of atomic and molecular hydrogen.[11] Equations (1)–(3) reduceat lto the classical "H_2 R-D" model [Eq. (3) of Krishnan *et al.*[3]] if k_H and k_{H_2} are so large that $H \leftrightarrow H_2$ transformation becomes instantaneous.

Our simulation results, based on the solution of Eqs. (1)–(3), show that the dominance of $H \leftrightarrow H_2$ conversion during short-term stress extends the transition between reaction-limited region ($n \sim 1$) and H_2 diffusion-limited region ($n \sim 1/6$) by several orders of time scale, enabling excellent fit for short-term NBTI degradation [see Fig. 1(a)]. Later when $H \leftrightarrow H_2$ conversion reaches quasiequilibrium, H_2 diffusion dominates and results in $n \sim 1/6$. We use the extensively verified field dependence of k_f (Ref. 6) to interpret the voltage-dependent data in Fig. 1. This figure also shows that the generalized "H–H_2 R-D" model *at long times* predicts the same time exponent as the "H_2-only R-D" model, because in steady state, diffusion of H_2 governs the dynamics of interface trap generation.[12] Slightly lower n, observed for experimental long t_{stress} data in Fig. 1(a), can be attributed to the decrease in (stress) electric field as a function of time. NBTI causes $|V_T|$ to increase, hence for a constant stress bias, the oxide electric field decreases. Our numerical calculation predicts a decrease in n to ~ 0.15 [typically observed for OTFM (Refs. 2, 5, and 8)] for maximum stress condition of Fig. 1(a). Moreover, as similar stress field dependence of k_f explains NBTI stress data at all stress times in Fig. 1(a), we recognize that for these devices, *any contributions to* ΔV_T *from hole trapping is negligible*. Further, since activation energies for hole trapping,[9] H and H_2 diffusion, and N_{IT} generation[2,3,6,13] (represented by $E_{A(HT)}$, $E_{A(H)}$, $E_{A(H_2)}$, $E_{A(IT)}$, respectively) are such that $E_{A(HT)} \ll E_{A(H)} < E_{A(H_2)} \sim E_{A(IT)}/n$, models involving hole trapping[9] would be inconsistent with temperature-dependent data for these specific experiments.

Next we address the puzzle of measurement-dependent n values for *short-time* degradation (see Fig. 2). Extracting the physics behind NBTI degradation at any stress time requires one to obtain the exact nature of ΔV_T using some experimental setup. Indeed classical I_D-V_G measurement provides an incorrect estimate of n at mid- to long stress time, due to unintentional recovery inherent in the measurement technique.[2,14] The uninterrupted OTFM eliminates the "recovery" problem[2,5,8,14–16] and calculates ΔV_T by: (a) $\Delta V_T(t) = |(I_{dlin}(t) - I_{dlin}(t_0))/I_{dlin}(t_0)|^* V_{GT0}$ (Refs. 2, 5, 8, and 14) (see Ref. 17 for details about the calculation done by Rangan *et al.*[14]), (b) $\Delta V_T(t) = \int_{t_0}^{t} \partial I_{d,sense}/g_{m,sense}$,[16] or (c) $\Delta V_T(t) = M|(I_{dlin}(t) - I_{dlin}(t_0))/I_{dlin}(t_0)|$.[15] These methods give comparable results for long-term NBTI stress.[15] A point that is not always obvious, hence ignored, is the use of t_0 in the above calculation. Exact ΔV_T measurement using OTFM mandates the use of $t_0 = 0$. Yet OTFM setups require some time (≥ 1 ms) before measuring $I_{dlin}(t_0)$; as a result $t_0 \neq 0$ (termed as time-zero delay) in such measurements. Therefore, while solving the issue of unintentional recovery at each measurement step, OTFM introduces an error in degradation estimates due to nonzero t_0, the effect of which will be particularly important at short stress times. Such t_0 error is not present either in classical I_D-V_G measurement (which is erroneous due to recovery issue, discussed earlier) or in the UFM, because in these cases $\Delta V_T(0)$ is obtained *before* the device is stressed. To observe the effect of time-zero delay in the OTFM results, we have used the experimental data from Reisinger *et al.*[9] (at $V_{G,stress} = -2.2$ V, 125 °C), with a specified measurement delay of ~ 1 μs (hence very small recovery) and $t_0 \sim 0$ (similar to classical I_D-V_G), as reference. Once correction due to t_0 is accounted for, we can explain n

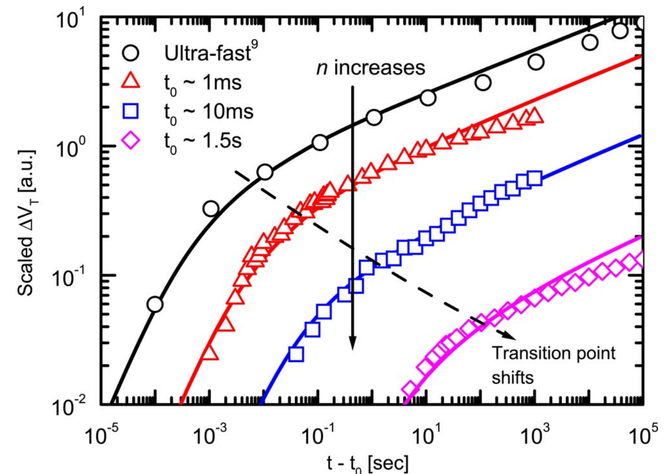

FIG. 2. (Color online) Effect of t_0 on OTFM compared with UFM, for similar stress conditions. Here, $\Delta V_T(t-t_0) \sim |(N_{IT}(t) - N_{IT}(t_0))/N_{IT}(t_0)|$ and data are scaled along the Y axis for clarity. As t_0 is increased, there is an increase in time exponent at a particular $(t-t_0)$ mainly in the short-term stress and also a rightward shift of the apparent transition point when the system reaches steady state with H_2 diffusion.

Downloaded 25 May 2007 to 128.46.94.118. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

for all OTFM data in Fig. 2, having various $t_0 \neq 0$. The figure indicates that the use of $t_0 \neq 0$ in calculating $\Delta V_T(t)$ for OTFM results in an "apparent and artificial" increase of n mainly in short-term NBTI degradation. Indeed, if t_0 is very high (>1 s), even the longer-term stress data may be effected by t_0 delay. This delay explains the discrepancy of OTFM results from different experimental groups[2,14,18] regarding the onset of the $n \sim 1/6$ regime (see Fig. 2). As consideration of t_0 delay alone can explain the variation of n in Fig. 2, any possible contribution from hole trapping must have fully saturated[9] before t_0 and hence will have no effect in plotted ΔV_T. Therefore, when devices have comparatively few hole trapping,[8] 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. And these t_0-corrected experimental data can then be interpreted by the $H-H_2$ model [Fig. 1(a)]. As a practical matter, if such t_0 correction of OTFM is difficult, our simulation suggests that reasonable projection for NBTI lifetime can still be made if t_0 is kept below 0.2 ms and projection is based on the data above 1 s. On the other hand if hole trapping is significant (as would be the case for films having thick thermal SiON dielectric[8]) and saturates before t_0, time-zero delay correction allows one to extract back *only* the component of the degradation due to interface traps. To determine whether hole trapping had saturated before t_0 in high $N_{\text{interface}}$ samples, one should make a second calculation/measurement with higher time-zero delay and check whether the same ΔV_T vs t (for $t_0=0$) can be simultaneously extracted back from both data sets.

In conclusion, we have illustrated how a generalized $H-H_2$ R-D model can consistently explain the NBTI degradation at both short- and long-term stresses. We trace the previous discrepancy between theory and measurements to the assumption of "instantaneous H_2 formation" in the classical model. Further, we demonstrate that the existing discrepancy in short-term NBTI data by various measurement techniques can be attributed to contamination by time-zero delay in OTFM systems. Our analysis suggests that lifetime prediction[4,6] and projection back to operating condition[6] must be carried out using long-term NBTI stress data (having $n \sim 1/6$), obtained after $t_{\text{stress}} \sim 100$ s with $t_0 \leqslant 1$ ms, because higher exponents in short-term degradation can result in unnecessarily pessimistic projections.

This work was done with financial support from Applied Materials, TSMC, and SRC. The authors also acknowledge Souvik Mahapatra for invaluable suggestions and the Network of Computational Nanotechnology (NCN) for providing computational resources.

[1] M. A. Alam and S. Mahapatra, Microelectron. Reliab. **45**, 71 (2005).
[2] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. A. Alam, Tech. Dig. - Int. Electron Devices Meet. **2005**, 684.
[3] A. T. Krishnan, S. Chakravarthi, P. Nicollian, V. Reddy, and S. Krishnan, Appl. Phys. Lett. **88**, 153518 (2006).
[4] C. L. Chen, Y. M. Lin, C. J. Wang, and K. Wu, Proceedings of IEEE IRPS, (IEEE, Piscataway, NJ, 2005), p. 704.
[5] G. Gupta, S. Mahapatra, L. Madhav, D. Varghese, K. Ahmed, and F. Nouri, Proceedings of IEEE IRPS, 2006 (IEEE, Piscataway, NJ, 2006), p. 731.
[6] A. E. Islam, G. Gupta, S. Mahapatra, A. Krishnan, K. Ahmed, F. Nouri, A. Oates, and M. A. Alam, Tech. Dig. - Int. Electron Devices Meet. **2006**, 329.
[7] S. Chakravarthi, A. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, Proceedings of IEEE IRPS (IEEE, Piscataway, NJ, 2004), p. 273.
[8] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, Proceedings of IEEE IRPS (IEEE, Piscataway, NJ, 2007).
[9] H. Reisinger, O. Blank, W. Heinrigs, A. Muhlhoff, W. Gustin, and C. Schlunder, Proceedings of IEEE IRPS (IEEE, Piscataway, NJ, 2006), p. 448.
[10] C. G. Van de Walle, Tech. Dig. - Int. Electron Devices Meet. **2005**, 400.
[11] H. Kufluoglu and M. A. Alam (unpublished).
[12] Work on analytical solution for $H-H_2$ model is under way.
[13] B. Tuttle, Phys. Rev. B **61**, 4417 (2000); C. G. Van de Walle and B. R. Tuttle, IEEE Trans. Electron Devices **47**, 1779 (2000).
[14] S. Rangan, N. Mielke, and E. C. C. Yeh, Tech. Dig. - Int. Electron Devices Meet. **2003**, 341.
[15] C. R. Parthasarathy, M. Denais, V. Huard, G. Ribes, E. Vincent, and A. Bravaix, Proceedings of IEEE IRPS (IEEE, Piscataway, NJ, 2006), p. 471.
[16] M. Denais, A. Bravaix, V. Huard, C. Parthasarathy, G. Ribes, F. Perrier, Y. Rey-Tauriac, and N. Revil, Tech. Dig. - Int. Electron Devices Meet. **2004**, 109.
[17] Although Rangan *et al.* (Ref. 14) has not done V_T calculation, their percentage degradation calculation uses a similar formula (within some scaling factor), considering negligible mobility degradation at high electric field (Ref. 15).
[18] V. Huard, C. R. Parthasarathy, C. Guerin, and M. Denais, Proceedings of IEEE IRPS (IEEE, Piscataway, NJ, 2006), p. 733.
[19] A. T. Krishnan, C. Chancellor, S. Chakravarthi, P. E. Nicollian, V. Reddy, A. Varghese, R. B. Khamankar, S. Krishnan, and L. Levitov, Tech. Dig. - Int. Electron Devices Meet. **2005**, 688.

Downloaded 25 May 2007 to 128.46.94.118. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp