Errors in Projecting Gate Dielectric Reliability From Fowler–Nordheim Stress to Direct-Tunneling Operation

Paul E. Nicollian, *Senior Member, IEEE*

***Abstract*—The extrapolation of gate dielectric reliability of devices that are stressed in the Fowler–Nordheim (F-N) tunneling regime but operate in the direct-tunneling (D-T) region is analyzed using the time-dependent dielectric-breakdown power law model. Due to the differences in the slopes of the gate current versus voltage characteristics between F-N and D-T, the widely established practice of directly extrapolating the time to breakdown from stress to operating conditions is, in general, not rigorously correct. Reliability projections across a wide voltage space can have significant errors, and the operating voltage that a technology can safely sustain may be overestimated.**

***Index Terms*—Breakdown, dielectric, oxide, reliability, SiON, time-dependent dielectric breakdown (TDDB).**

I. Introduction

It is widely known that the breakdown of gate dielectrics used to form the insulating films of MOSFET devices is a serious reliability concern. Models used to project the time to breakdown (t_{BD}) from accelerated stress down to lower voltage operating conditions have been extensively researched since the 1970s. In modern technologies with ultrathin gate dielectrics, the time-dependent dielectric-breakdown (TDDB) power law model [1] is commonly used for reliability assessments, where t_{BD} and the charge to breakdown (Q_{BD}) follow a power law in gate voltage (V_G)

$$t_{\mathrm{BD}} = a_T V_G^{-N_T} \tag{1}$$

$$Q_{\mathrm{BD}} = a_Q V_G^{-N_Q}. \tag{2}$$

The power law model arises when the mechanism for the generation of trap states that lead to breakdown is the vibrational excitation of silicon hydrogen bonds [2]. Since the exponents N_T and N_Q in (1) and (2) can be greater than 40 [1], [2], the TDDB power law model results in optimistic lifetimes compared to exponential models [1]. Accordingly, it should be carefully examined for any application for which it is utilized.

For polysilicon gate devices with $\mathrm{SiO_2}$ dielectrics, if the oxide voltage (V_{OX}) is greater than the approximately 3 V barrier height (Φ_B) for the Si-$\mathrm{SiO_2}$ interface, electrons can be injected from the cathode into the oxide conduction band via Fowler–Nordheim (F-N) tunneling through a triangular barrier [3], [4], as shown in Fig. 1(a) and (b). For F-N, the tunneling distance (X_T) decreases with increasing V_{OX}. However, if the dielectric thickness (t_{OX}) is less than about 4 nm and $V_{\mathrm{OX}} < \Phi_B$, then electrons can direct tunnel through a trapezoidal barrier [5] from the cathode to the anode without entering the oxide conduction band, as shown in Fig. 1(c) and (d). Unlike F-N tunneling, for direct tunneling (D-T), X_T is independent of V_{OX} and is always equal to t_{OX}. For this reason, the slope of the current–voltage curve will be higher for F-N compared to D-T, as shown in Fig. 1(e) and (f). For oxides where nitrogen has been incorporated to form SiON, Φ_B can be less than 3 V [6], [7]. Accordingly, TDDB stress of ultrathin SiON films may occur in the F-N region. In this letter, we will show that the higher F-N slope can lead to erroneous reliability projections if accelerated testing is performed in the F-N regime and the device operates in the D-T region. Consequently, the widely used approach of directly extrapolating t_{BD} to operating conditions [1], [8], [9] is, in general, not valid. We show a simple correction method to obtain accurate lifetimes.

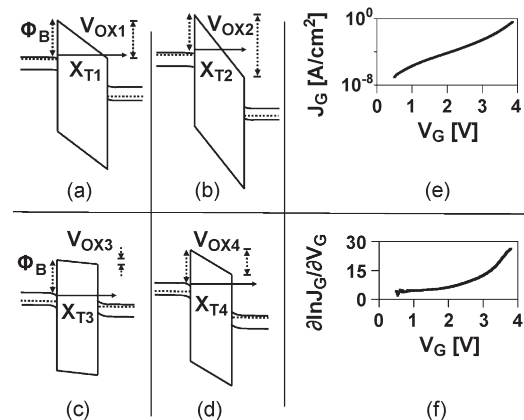

Fig. 1. Band diagrams for [(a) and (b)] F-N and [(c) and (d)] D-T. For F-N, when $V_{\mathrm{OX2}} > V_{\mathrm{OX1}}$, $X_{T2} < X_{T1}$. For D-T, when $V_{\mathrm{OX4}} > V_{\mathrm{OX3}}$, $X_{T4} = X_{T3}$. (e) Experimental J_G versus V_G curve for a 3.0 nm film and (f) the corresponding slope $\partial \ln(J_G)/\partial V_G$ showing the effects of the transition from D-T to F-N.

II. Results

Traps are generated in the oxide bulk and at its contact interfaces when a voltage is applied across the gate dielectric of a MOS device and a current passes through the film [10]. The buildup of these traps leads to the breakdown of the dielectric.

Manuscript received June 23, 2009; revised July 22, 2009. First published October 13, 2009; current version published October 23, 2009. This work was supported by Texas Instruments Incorporated. The review of this letter was arranged by Editor B.-G. Park.

The author is with Texas Instruments Incorporated, Dallas, TX 75243 USA (e-mail: nicollian@ti.com).

Digital Object Identifier 10.1109/LED.2009.2030698

0741-3106/$26.00 © 2009 IEEE

Authorized licensed use limited to: TEXAS INSTRUMENTS VIRTUAL LIBRARY. Downloaded on December 16, 2009 at 19:52 from IEEE Xplore. Restrictions apply.

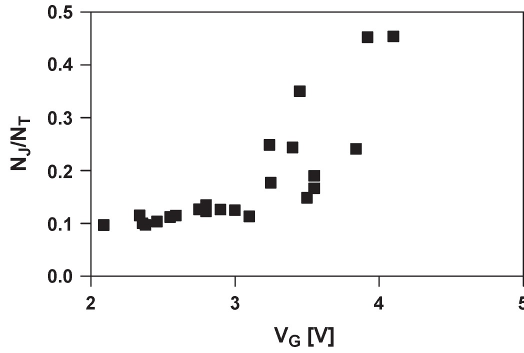

Fig. 2. Ratio of N_J/N_T versus V_G. N_J/N_T becomes strongly V_G dependent and of nonnegligible magnitude when devices are stressed in the F-N regime. The EOTs of the devices in this plot range from 1.0 to 3.2 nm.

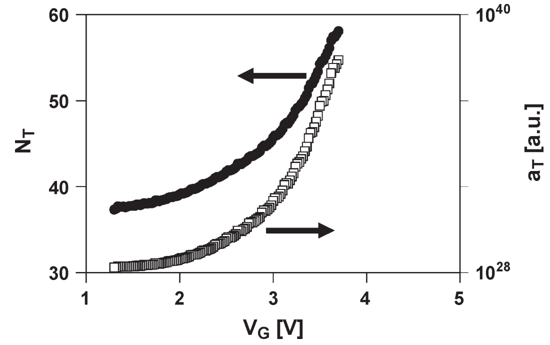

Fig. 3. TDDB power law model exponent ($y1$-axis) and prefactor ($y2$-axis) versus V_G. The model parameters in (1) become V_G dependent when the voltage dependence of tunneling has been accounted for. The EOT is 3.0 nm, and the devices were stressed in the F-N regime.

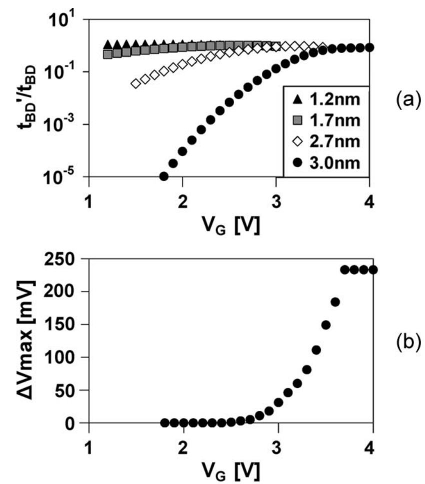

Fig. 4. (a) Ratio of $t_{\rm BD}'/t_{\rm BD}$ for corrected to uncorrected TDDB model parameters versus V_G. The lifetime at operating conditions is overly optimistic by several orders of magnitude for F-N stress and D-T operation. (b) Change in safe operating voltage $\Delta V_{\rm MAX}$ for uncorrected models relative to corrected models. The EOT is 3.0 nm, and the devices were stressed in the F-N regime. For this case, the maximum safe operating voltage can be overestimated by up to 250 mV for F-N stress and D-T operation.

The number of generated traps $N(Q)$ increases as a power law function of fluence [11]

$$N(Q) = bQ^m. \quad (3)$$

At breakdown, $N(Q) = N_{\rm BD}$ and $Q = Q_{\rm BD}$. Rearrangement of (3) yields

$$Q_{\rm BD} = (N_{\rm BD}/b)^{1/m}. \quad (4)$$

The trap generation rate (P_G) is calculated as

$$P_G = \partial N(Q)/\partial Q = mbQ_{\rm BD}^{m-1}. \quad (5)$$

Solving (4) for b and inserting it into (5) and rearranging terms yields

$$Q_{\rm BD} = mN_{\rm BD}/P_G. \quad (6)$$

The relationship between $t_{\rm BD}$ and $Q_{\rm BD}$ is through the current density (J_G)

$$t_{\rm BD} = Q_{\rm BD}/J_G. \quad (7)$$

The only V_G-dependent term for $Q_{\rm BD}$ is P_G, whereas $t_{\rm BD}$ is V_G dependent through P_G and J_G. The power law model exponents N_T, N_Q, and N_J defined for time, fluence, and current, respectively, are

$$N_T = -\,\partial(\ln t_{\rm BD})/\partial \ln(V_G) \quad (8)$$

$$N_Q = -\,\partial(\ln Q_{\rm BD})/\partial \ln(V_G) \quad (9)$$

$$N_J = \partial(\ln J_G)/\partial \ln(V_G). \quad (10)$$

Inserting (8)–(10) into (7) yields

$$N_T = N_Q + N_J. \quad (11)$$

For (11) to be valid, $\ln(t_{\rm BD})$ versus $\ln(V_G)$ must be a straight line. However, N_J is not a constant slope due to the transition from F-N to D-T at $V_{\rm OX} = \Phi_B$. Moreover, J_G is not a power law in V_G [3]–[5] as required in (10). For these reasons, strictly speaking, $t_{\rm BD}$ is not a power law in V_G, so the direct application of (1) is not rigorously correct.

The ratio N_J/N_T for the first breakdown event is shown in Fig. 2 for NMOS devices with plasma-nitrided oxide films, with 25 different equivalent oxide thicknesses (EOTs) ranging from 1.0 to 3.2 nm. This variation of EOT enables the evaluation of N_J/N_T over a large voltage range. The contribution from the current slope term N_J approaches 50% of N_T for F-N stress, whereas it only comprises 10% of N_T in the D-T region. This artifact is eliminated by extrapolating $Q_{\rm BD}$ from stress to lower voltages using (2). After $Q_{\rm BD}$ is converted to $t_{\rm BD}$ using (7), a_T and N_T are recalculated iteratively from least squares fits using (1). This eliminates the erroneously high values of these parameters resulting from the N_J term.

The TDDB power law model parameters corrected for N_J are shown in Fig. 3 for 3.0 nm EOT films stressed in the F-N regime. Scaling from 3.7 V stress to 1.8 V operation, N_T decreases from 58 to 38, while a_T drops by 9 orders of magnitude. The effect on $t_{\rm BD}$ is shown in Fig. 4(a), where $t_{\rm BD}$ is the lifetime extrapolated directly from (1) and $t_{\rm BD}'$ is the lifetime from corrected parameters. For the thicker films that were subjected to F-N stress, lifetimes for uncorrected models can be overly optimistic by several orders of magnitude, whereas there is little impact on devices that were stressed in the D-T region. For a five-fit average failure rate requirement, the

Authorized licensed use limited to: TEXAS INSTRUMENTS VIRTUAL LIBRARY. Downloaded on December 16, 2009 at 19:52 from IEEE Xplore. Restrictions apply.

maximum safe operating voltage for the 3.0 nm films that were subjected to F-N stress can be overestimated by nearly 250 mV, as shown in Fig. 4(b). In obtaining this result, we have assumed that N_Q is independent of V_G. The cooperation between vibrational and electronic excitation modes of the silicon–hydrogen bond, which can result in N_Q being lower at accelerated stress compared to operating conditions [2], has not been factored in because this potential gain may be offset by detrimental effects of the nitrogen profile such as nonuniformity [9].

Whereas correcting for the differences in the voltage dependences of F-N and D-T in SiON films results in lower projected lifetimes at operating conditions, transport can have the opposite effect on the reliability of high-k stacks. In the SiO_2/ZrO_2 system, accounting for J_G results in *higher* projected lifetimes for substrate injection due to the ZrO_2 layer becoming a tunneling barrier at low voltages [12]. This is a different effect than we report in this letter, as it will occur whether transport in the SiO_2 portion of the SiO_2/ZrO_2 stack is D-T or F-N. Accordingly, the effect of J_G on TDDB must be separately evaluated for any gate stack of interest.

III. Conclusion

For devices that are stressed in the F-N tunneling regime but operate in the D-T region, dielectric reliability predictions are unrealistically optimistic if projections using only the time to breakdown are applied. This artifact arises from the differences in the slopes of the current–voltage characteristics between F-N and D-T. This problem is circumvented by invoking Q_{BD} in projecting breakdown from stress to operating conditions. Accordingly, the widely used approach of extrapolating dielectric reliability directly from established time to breakdown models is, in general, not correct.

Acknowledgment

The author would like to thank J. Ondrusek and S. Krishnan for their support of this letter.

References

[1] E. Y. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L.-K. Han, C. Montrose, and R. Dufresne, "Voltage-dependent voltage-acceleration of oxide breakdown for ultra-thin oxides," in *IEDM Tech. Dig.*, 2000, pp. 541–544.

[2] J. Suñé and E. Y. Wu, "Mechanisms of hydrogen release in the breakdown of SiO_2-based gate oxides," in *IEDM Tech. Dig.*, 2005, pp. 399–402.

[3] R. H. Fowler and L. Nordheim, "Electron emission in intense electric fields," *Proc. R. Soc. Lond. A*, vol. 119, no. 781, pp. 173–181, 1928.

[4] E. H. Snow, "Fowler–Nordheim tunneling in SiO_2 films," *Solid State Commun.*, vol. 5, no. 10, pp. 813–815, 1967.

[5] K. F. Schuegraf, C. C. King, and C. Hu, "Ultra thin silicon dioxide leakage current and scaling limit," in *VLSI Symp. Tech. Dig.*, 1992, pp. 18–19.

[6] D. M. Brown, P. V. Gray, F. K. Heumann, H. R. Philipp, and E. A. Taft, "Properties of $Si_XO_YN_Z$ films on Si," *J. Electrochem. Soc.*, vol. 115, no. 3, pp. 311–317, 1968.

[7] X. Guo and T. P. Ma, "Tunneling leakage current in oxynitride: Dependence on oxygen/nitrogen content," *IEEE Electron Device Lett.*, vol. 19, no. 6, pp. 207–209, Jun. 1998.

[8] R. Degraeve, N. Pangon, B. Kaczer, T. Nigam, G. Groeseneken, and A. Naem, "Temperature acceleration of oxide breakdown and its impact on ultra thin gate oxide reliability," in *VLSI Symp. Tech. Dig.*, 1999, pp. 59–60.

[9] P. E. Nicollian, G. C. Baldwin, K. N. Eason, D. T. Grider, S. V. Hattangady, J. C. Hu, W. R. Hunter, M. Rodder, and A. L. P. Rotondaro, "Extending the reliability scaling limit of SiO_2 through plasma nitridation," in *IEDM Tech. Dig.*, 2000, pp. 545–548.

[10] D. J. DiMaria, "Explanation for the polarity dependence of breakdown in ultra thin silicon dioxide films," *Appl. Phys. Lett.*, vol. 68, no. 21, pp. 3004–3006, 1996.

[11] D. J. DiMaria and E. Cartier, "Mechanism for stress induced leakage currents in thin silicon dioxide films," *J. Appl. Phys.*, vol. 78, no. 6, pp. 3883–3894, 1995.

[12] T. Kauerauf, R. Degraeve, E. Cartier, B. Govoreanu, P. Blomme, B. Kaczer, L. Pantisano, A. Kerber, and G. Groeseneken, "Towards understanding degradation and breakdown of SiO_2/high-k stacks," in *IEDM Tech. Dig.*, 2002, pp. 521–524.

Authorized licensed use limited to: TEXAS INSTRUMENTS VIRTUAL LIBRARY. Downloaded on December 16, 2009 at 19:52 from IEEE Xplore. Restrictions apply.