Charge-to-Breakdown and Trap Generation Process in Thin Oxides

Gennadi Bersuker, James Werking, and David Y. Chan
SEMATECH, 2706 Montopolis Dr., Austin, TX 78741
Ph.:512-356-7045; gennadi.bersuker@sematech.org

Abstract—In the proposed model, trap generation is assumed to be triggered by the collision of injected electrons with oxide atoms. The model suggests that thinner oxides are less susceptible to charging stress due to both lower probability of electron collision and lower electron impact energy. The difference in positive and negative gate bias charge-to-breakdown data is attributed to the formation of the structural transition layer at the Si/SiO_2 interface. The model is used for analysis of the effects of process induced charging damage on transistor parameters. It is found that after heavy stress, leakage current is determined by the probability of trap assisted tunneling, while the density of generated traps controls leakage in lightly damaged oxides.

Introduction

Scaled devices require thinner gate oxides, and it is therefore important to estimate how such scaling affects oxide susceptibility to process induced charging damage. To address this issue, we used test structures that contain transistors with attached charge collecting antennas at different wafer processing levels. The sensitivity of transistor parameters to gate oxide characteristics allows for measurement of relatively low damage that may not show up in yield but can affect device reliability.

Complementary information can be obtained from charge-to-breakdown, $\mathbf{Q_{bd}}$, measurements that can emulate to some extent charging stress during plasma process steps. However, interpretation of $\mathbf{Q_{bd}}$ data is not straightforward since the nature of stress induced traps is not always known, and may vary with stress conditions. Therefore, there is a need for a model that can correlate, at least phenomenologically, damage in the oxides with oxide characteristics and electrical stress conditions.

Model For Trap Generation Process

We propose a model for the trap generation process that can be applied to ultra thin oxides. The central point of this model is an assumption that whatever the nature of the oxide damage induced by electrical stress, this damage is triggered by the collision of injected electrons with oxide atoms (which seems to be a dominant process in thin oxides [1] under low bias conditions [2]). If we accept the idea that no damage is possible without initial interaction between the injected electrons and the lattice, then number of generated traps per unit area can be written as follows

$$\mathbf{D = Q \cdot n}, \qquad (1)$$

where **Q** is total number of electrons per unit area passed through the oxide, and **n** is the probability of trap creation per one electron (analog of the absorption coefficient): $\mathbf{n = 1 - exp[-p \cdot s \cdot (t - t_0)]}$. Here **s** is the electron collision cross section, **p** is the probability of trap creation in the result of a collision, **t** and $\mathbf{t_0}$ are oxide thickness and electron direct tunneling thickness, respectively. Due to low probability of trap creation [3], it is assumed that each electron has only one chance to create a trap, and also, the magnitude of **n** for the direct tunneling process is negligible. **p** increases with the electron energy $\mathbf{E_{el}}$ and decreases with the bond strength: $\mathbf{p = p_0 E_{el} = p_0 E_{ox} t}$, where $\mathbf{E_{ox}}$ is the electric field in the oxide, and $\mathbf{p_0}$ is the temperature dependent probability to break the bond in SiO_2. Oxide thickness is assumed to be comparable to the electron mean free path.

Thus, with the constant current stress condition, the dependence of trap density on oxide thickness for ultra thin oxides can be written as follows:

$$\mathbf{D \cong Q \cdot s \cdot p_0 \cdot E_{ox} \cdot (t - t_0)} \qquad (2)$$

Based on Eq. (2), one may expect less traps to be generated in thinner oxides.

Oxide breakdown occurs when local trap density exceeds some characteristic critical value $\mathbf{D_c}$. Taking into account approximately linear dependence of critical trap density $\mathbf{D_c}$ on oxide thickness in very thin oxides [4], $\mathbf{D_c \cong D_{c0} \cdot t}$ ($\mathbf{D_{c0}}$ is the critical concentration in the very thin oxide $\mathbf{t \cong t_0}$), from Eq. (2) the following dependence of charge-to-breakdown $\mathbf{Q_{bd}}$ on oxide thickness is obtained:

$$\mathbf{Q_{bd} = (D_{c0}/s \cdot p_0 \cdot E_{ox}) \cdot (t/t - t_0)} \qquad (3)$$

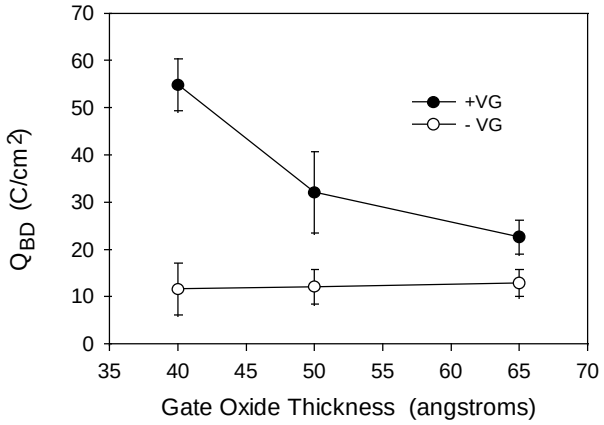

Fig.1. Charge-to-breakdown vs. oxide thickness.

In order to verify the above $\mathbf{Q_{bd}(t)}$ dependence, we compared oxide thickness **t** calculated with Eq. (3) to the nominal thickness that was measured both optically and electrically. Charge-to-breakdown measurements were done in 20x20 um^2 capacitors under positive and negative gate bias constant current stress, Fig. 1.

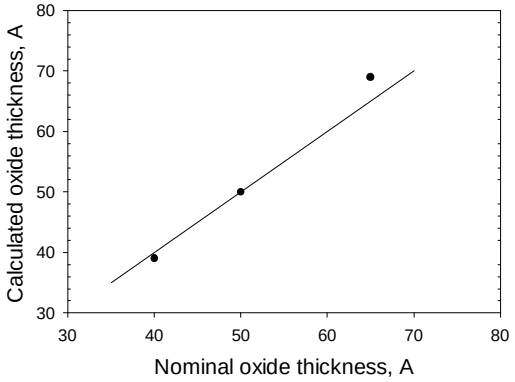

Fig. 2. Gate oxide thickness calculated with Eq. (3) vs. nominal thickness.

0-7803-4205-4/97/$10.00 ©1997 IEEE

Using positive bias $\mathbf{Q_{bd}}$ data and ratios of $\mathbf{Q_{bd}}$ in Eq. (3) for different **t**, we found that the best fit of the calculated oxide thickness to the nominal ones, Fig. 2, can be obtained with $\mathbf{t_0} \cong 30$ Å.

So far, the oxide parameters **s** and $\mathbf{p_0}$ have been assumed to remain constant across the oxide. However, it is well established that stress at the oxide/Si interface results in a structural transition layer in SiO_2 that may extend up to few tens of Å from the interface. The transition layer is characterized by higher than bulk oxide density (2.4 g/cm^3) and high compressive stress [5]. These lead to increase in both the electron collision cross section, **s,** and the probability of bond breaking, $\mathbf{p_0}$, in the transition layer.

The existence of the transition layer breaks the symmetry between the positive and negative gate bias injections. Indeed, since in the case of substrate electron injection direct tunneling distance $\mathbf{t_0} \cong 30$ Å
covers significant portion of the transition layer, **s** and $\mathbf{p_0}$ values may be expected to be very close to the bulk oxide ones, Fig. 3 .

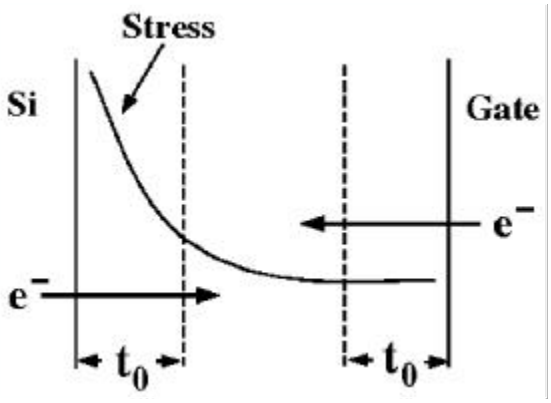

Fig. 3 Substrate and gate electron injections vs. distribution of stress in oxides.

On the other hand, for gate injection, the direct tunneling distance covers oxide outside the structural transition layer. Therefore, to a great extent, **s** and $\mathbf{p_0}$ values in Eq. (3), which represent averages of $\mathbf{s(t)}$ and $\mathbf{p_0(t)}$ over the oxide thickness **t** outside the direct tunneling distance, are determined by the structural transition layer properties, and thus are much greater than in the substrate injection case. Subsequently, in the latter case, $\mathbf{Q_{bd}}$ values are greater for all oxide thickness, as follows from Eq. (3). In addition, in the gate injection case, **s** and $\mathbf{p_0}$ parameters may be expected to exhibit strong oxide thickness dependence: the thinner oxide is, the greater the share of the structural transition layer in the total oxide thickness, and therefore average **s** and $\mathbf{p_0}$ values are bigger. Based on expected behavior of **s** and $\mathbf{p_0}$ parameters in the gate injection case in the extremely thin and thick oxides, we may propose the following dependence of the product $\mathbf{(s\cdot p_0)_g}$ on oxide thickness: $\mathbf{(s\cdot p_0)_g \sim (s\cdot p_0)_s\, t/t\text{-}t_0}$, where the product $\mathbf{(s\cdot p_0)_s}$, in the substrate injection case is independent of oxide thickness. Substituting the above $\mathbf{(s\cdot p_0)_g}$ dependence in the Eq. (3), we obtain that the following relation between the charge-to-breakdown for gate, $\mathbf{Q_{bdg}}$, and substrate, $\mathbf{Q_{bds}}$, injections valid for any oxide thickness;

$$\mathbf{Q_{bds}/Q_{bdg} = (s\cdot p_0)_g/(s\cdot p_0)_s \sim t/t\text{-}t_0} \qquad (4)$$

We calculated the oxide thickness **t** in Eq. (4) using $\mathbf{Q_{bd}}$ values in Fig. 1 and direct tunneling distance $\mathbf{t_0}$ = 30 Å. Comparison of the calculated thickness to the nominal ones is presented in Fig. 4. When oxide thickness is reduced below the direct tunneling thickness, $\mathbf{Q_{bdg}}$ may be expected to increase and eventually become equal to $\mathbf{Q_{bds}}$. Note that other oxide properties, like interface roughness may also contribute to $\mathbf{Q_{bd}}$ values.

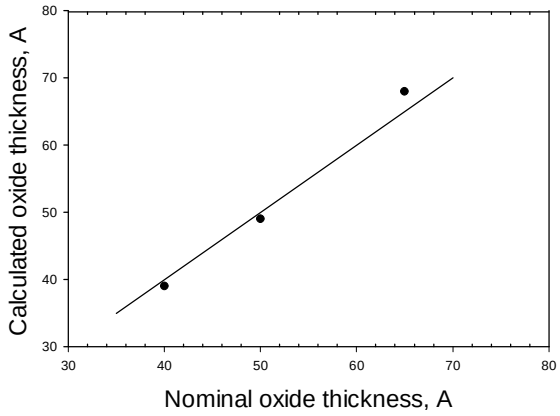

Fig. 4. Gate oxide thickness calculated with Eq. (4) vs. nominal oxide thickness.

PROCESS INDUCED DAMAGE

The above model was verified using process induced damage data. We analyzed charging damage effects in 0.35 um LDD NMOS transistors with 65 Å, 50 Å, and 40 Å gate oxides (similar effects were observed in the identically processed lot with 65 Å, 55 Å and 45 Å gate oxide splits). Splits for each oxide thickness contain 4 wafers, with 17 dies per wafer tested. Poly, metal-1 or full flow charge collecting antennas are attached to transistor gates, with antenna-to-gate ratios (AR) up to 90 K:1. A typical test tructure is shown in Fig. 5. Fowler-Nordheim (F-N) stress, 1 nA/um^2 for 2 sec, was used to reveal latent damage.

The charge damage effect (CDE) induced by processing is measured as a difference in parameter P shift due to F-N stress between an antenna transistor and a reference transistor at the same die location:

$$\mathbf{\Delta P = [P_{stress}(ant) - P_{nostress}(ant)] - [P_{stress}(ref) - P_{nostress}(ref)]} \qquad (5)$$

Reference devices are protected from charging damage during processing by fuses between drain, source, gate, and substrate.

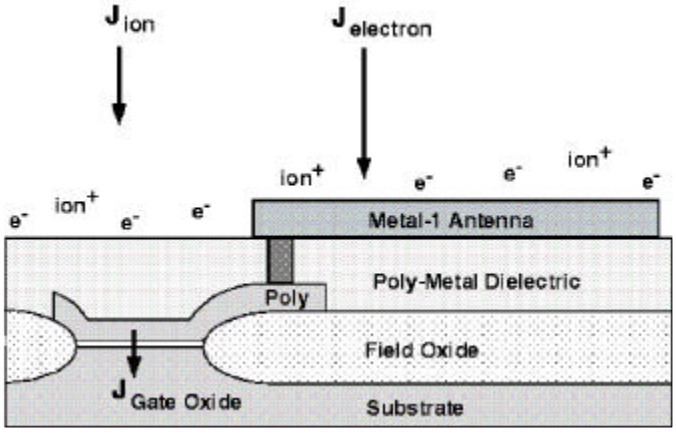

Fig. 5. Schematic presentation of charge collecting metal-1 antenna transistor after metal-1 etch.

In the case of thin oxide, gate leakage current I_g seems to be a very sensitive damage parameter; it exhibits strong antenna effect in most types of antenna modules even before F-N stress. In metal-1

antenna transistors, I_g increases with antenna size, Fig.6, and decreases with oxide thickness, Fig.7. The same trends are observed in CDE of I_g, after the F-N stress, Fig.8. Transconductance G_m and threshold voltage V_t do not show antenna effect before F-N stress.

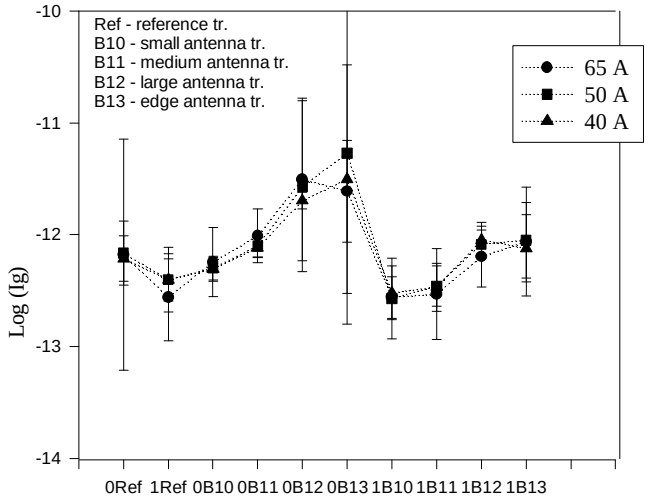

Fig. 6. Pre- and post-stress gate leakage current for reference and metal-1 antenna transistors. Error bars correspond to 1 standard deviation. In the x axis, 0 and 1 in front of the transistor type correspond to the data before and after the F-N stress.

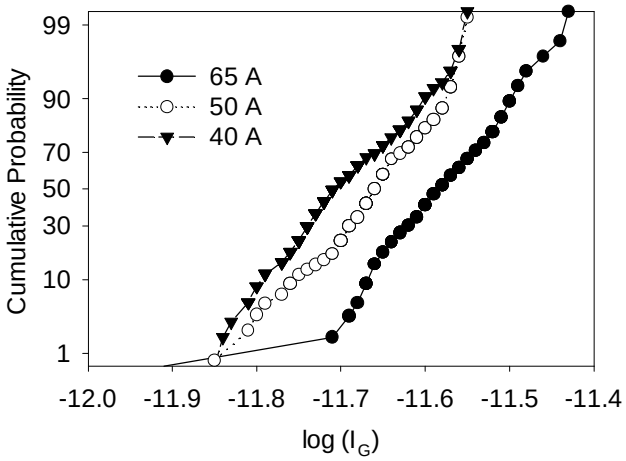

Fig. 7. Pre-stress gate leakage current of metal-1 large antenna transistors.

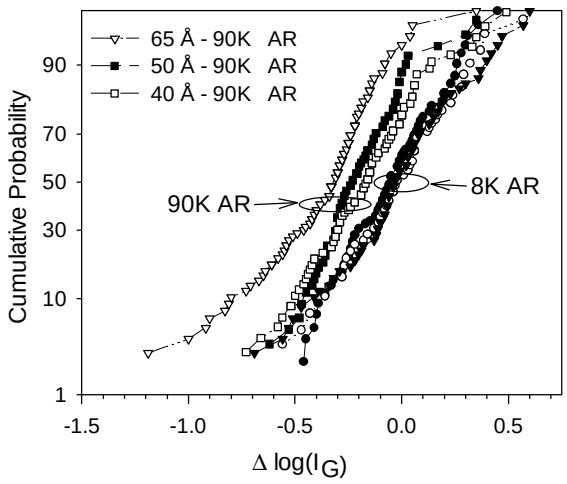

Fig. 8. Charge damage effect, calculated with Eq. (5), in gate leakage current of metal-1 antenna transistors

CDE in G_m and V_t is greater for thicker oxide, especially in devices with edge intensive antennas, Fig.9.

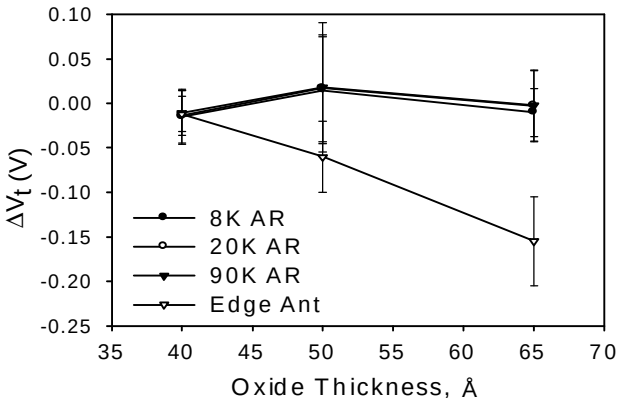

Fig. 9. Charge damage effect in threshold voltage of metal-1 antenna transistors.

The fact that I_g is reduced after the F-N stress, Fig.6, indicates that hole-type traps were created in the oxide during metal-1 processing. This conclusion is supported by V_t data, Fig. 9, that show negative shift of V_t with respect to reference transistors after F-N stress, such shift being accompanied by increase of G_m. Devices with poly area antennas show positive CDE of V_t, Fig.10, and no effect in I_g, suggesting that electron-type traps were primarily generated during poly processing.

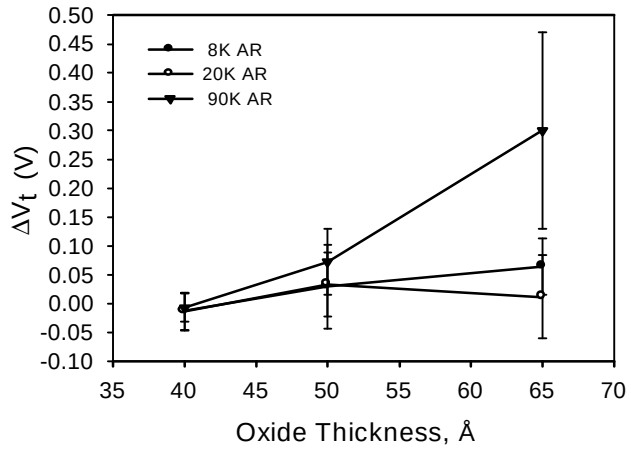

Fig. 10. Charge damage effect in threshold voltage of poly antenna transistors.

For tested transistor parameters in metal-1 and poly antenna modules, thicker gate oxides demonstrate greater susceptibility to charging stress, Figs. 6-10, (see [6]), in full agreement with the model. On the other hand, in the full flow modules that include contact antennas (having up to 39,000 contacts) in addition to area antennas, thinner oxides show higher gate leakage, Fig.11, (similar results were reported in [7]).

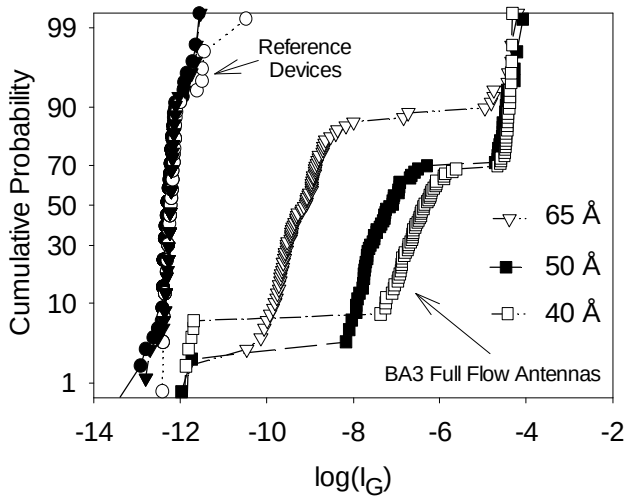

Fig. 11. Gate leakage current of reference and full flow large antenna transistors.

In this module, gate oxides in antenna transistors are very leaky due to heavy charging damage during processing, as follows from the dependence of I_g on antenna sizes. V_t and G_m show no antenna effect before F-N stress (no data available after the stress due to very high damage in transistors).

Thus, two groups of devices, with lower (metal-1, poly modules) and higher (full flow module) charging damage, demonstrate opposite trends: when
total damage is low, thinner oxides show less charge damage effect, however, they seem to be more susceptible to stress in the case of higher damage.

Stress Induced Leakage Current

In order to explain data in the heavily damaged full flow module that seemingly contradicts our model, we need to discuss a tional to the number of traps **D** which can accept electrons and to the probability of electron tunneling from the substrate to these traps **P** (for simplicity, we assume that in the ultra thin oxides, tunneling via one trap only is required for each electron):

$$\mathbf{j_{TA} = j_0\, D\, P} \qquad (6)$$

In the quasi-classical approximation, electron tunneling probability **P** can be written as follows:

$$\mathbf{P \propto exp\,\{-\int\sqrt{[E(t)-E_0]}dt\} \approx exp\,\{-E_b^{1/2}(\alpha+\beta t)\}},$$

where $\mathbf{E_b}$ and $\mathbf{E_0}$ are the energy barrier at the Si/SiO_2 interface and the initial electron energy, respectively. α and β are coefficients that depend on the electron and barrier energies and on the trap location between the oxide interfaces. In particular, if a trap is in the middle of the SiO_2 layer, $\beta = 1/2$.

As oxide thickness increases, trap assisted tunneling current $\mathbf{j_{TA}}$ quickly reduces due to exponential decline of tunneling probability **P**. However, $\mathbf{j_{TA}}$ may also increase with oxide thickness because of increase in the trap concentration **D**, Eq. (2). To determine dependence of the trap assisted tunneling current on the oxide thickness, we estimated an oxide thickness $\mathbf{t_{max}}$ at which trap assisted current is maximum.

Graphical results of these calculations are presented in Fig. 12. As follows from the equation $\mathbf{dj_{TA}/dt = 0}$, $\mathbf{t_{max}}$ decreases with the increase of the trap concentration per unit oxide thickness $\mathbf{D_0 = Q\cdot s\cdot p_0 \cdot E_{ox}}$, $\mathbf{t_{max} \propto \gamma/D_0}$, where parameter γ is a function of $\mathbf{E_b}$, β, and $\mathbf{t_0}$. The value of $\mathbf{t_{max}}$ with respect to the oxide thickness under consideration tells us whether leakage current should increase or decrease with oxide thickness. When the trap concentration $\mathbf{D_0}$ is small, $\mathbf{t_{max}}$ may be greater than the maximum thickness of the oxides under consideration, and therefore one observes increase of leakage current as oxide thickness increases (broken line in Fig. 12), as we see in metal-1 and poly modules. Decline of current in thinner oxides is due to insufficient number of traps which assist tunneling current, and we term this case damage controlled leakage.

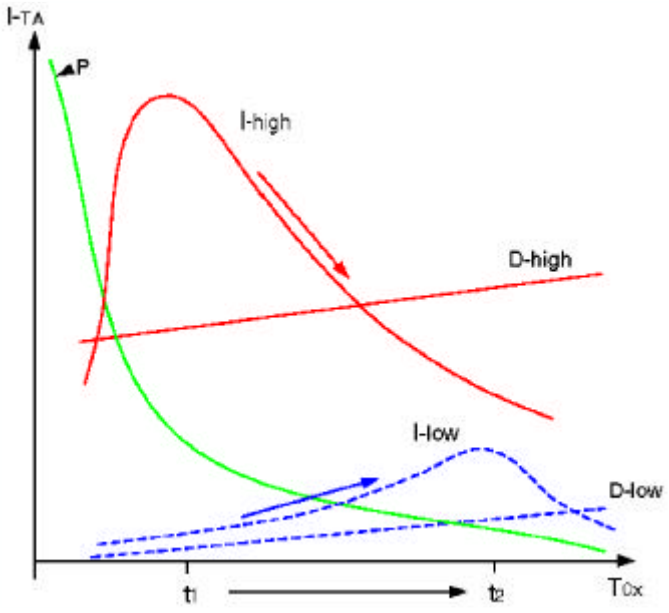

Fig. 12. Leakage current vs. oxide thickness in the cases of high, D-high, and low, D-low, damage.

When damage $\mathbf{D_0}$ increases as in the case of full flow module, $\mathbf{t_{max}}$ decreases and may correspond to oxide thickness that is smaller than the minimum thickness of the oxides in our experiment. In such case that can be termed tunneling controlled leakage, one observes decrease of leakage current when oxide thickness increases (solid line in Fig. 12), due to smaller tunneling probability **P** in thicker oxides.

Conclusion

With the same quality of oxides - effective scattering cross section **s** (determined by the concentration of charge defects, neutral traps, etc. in the native oxide) and bond strength $\mathbf{p_0}$ (determined by stress, H-bonds, etc.) - thinner oxides show greater $\mathbf{Q_{bd}}$ due to the geometrical factor. Therefore, to compare quality of oxides having different thickness, measured $\mathbf{Q_{bd}}$ values should be multiplied by the scaling factor $\mathbf{(t-t_0/t)}$.

Thinner oxides are less susceptible to charging stress since both the probability of collisions of injected electrons with oxide atoms and electron impact energy decrease with oxide thickness in the case of ultra thin oxides. However, relatively smaller damage (determined as a concentration of generated traps) may produce a greater effect in device performance, in particular, in gate oxide leakage current.

After heavy stress, leakage is mostly controlled by the probability of trap assisted tunneling (which is higher in thinner oxides). In lightly damaged oxides, leakage current is determined by the concentration of generated traps (which is less in thinner oxides).

Questions and Answers

Question: What is applicability of your model?
Answer: The upper limit of the oxide thickness is determined by the condition that oxide thickness should be comparable to the electron mean free path. Practically, it means less than two-three mean free paths. Different techniques give various estimates for it; a number around 34 Å was reported by several groups. In this case, the upper limit is around 70 Å.

Question: To simplify your expression for trap density, you suggested that the probability of trap creation per one electron is small. Do you have data to support this assumption?
Answer: There are several publications on this topic. In particular, S. K. Lai and R. Degraeve (Ref. 3 and 4) demonstrated that only a few percent of injected electrons generate traps and interface states.

Question: Critical trap density - can we measure or simulate it ?
Answer: It was simulated and, also, calculated based on measurements (see Ref. 4).

Question: Do you have hard numbers for stress gradient in the oxide ?
Answer: I do not have any numbers for stress magnitude. Oxide density near the interface was reported to be 2.4 g/cm^3 which is significantly higher than in bulk oxide and indicates high stress in that area.

Question: How does trap generation in your model depend on electron energy ?
Answer: Higher energy of injected electrons increases probability to break the bond and create a trap. Electrons are assumed to move in a ballistic regime.

Question: For the high damage case - you only show gate leakage data - do you see the same trend in the threshold voltage data?
Answer: No transistor characteristics were available after the F-N stress since this stress killed already heavily damaged devices.

Question: You use the model to explain the leakage current behavior - does the same model explain the threshold voltage behavior?
Answer: In this presentation we discussed data primarily on leakage current since it seems to be one of the most sensitive parameters. All measured transistor parameters, like transconductance, saturation current, etc., satisfy the suggested model.

Acknowledgments

Authors are grateful to Chand R. Viswanathan and Tomasz Brozek (UCLA), and Stephen J. Fonash and Osama O. Awadelkarim (Penn State) for stimulating discussions.

References

[1]. C. Felsch and E. Rosenbaum, Proc. Int. Rel. Phys. Symp., p. 142, 1995
[2]. D. J. DiMaria, IEEE Electron Dev. Lett., v. 16, p. 184, 1995
[3]. S.K. Lai, J. Appl. Phys., v. 54, p. 2540, 1983
[4]. R. Degraeve et al., IEDM, p. 863, 1995
[5]. A. Ishitani, Abstracts of 43rd AVS National Symposium, p. 25, 1996
[6]. A. B. Joshi et al., Proc. Int. Rel. Phys. Symp., p. 300, 1996
[7]. S. Kriahnan et al., IEDM, p. 315, 1995