On the “intrinsic” breakdown of thick gate oxide

Kin P Cheung
National Institute of Standards & Technology, Gaithersburg, MD U.S.A.

Abstract – The thick gate oxide breakdown mechanism has become an important topic again due to the rising demand for power electronics. The failure of the percolation model in explaining the observed Weibull shape factor, \( \beta \), seriously hampers the establishment of thick gate oxide breakdown models and the ability to project reliability from measurement data. In this work, lifetime shortening by oxide defects are simulated to produce degraded breakdown distributions that match experimentally observed \( \beta \). The result shows that even a low density of defects with the right energy is enough to greatly degrade \( \beta \) for thick oxides. Strong area scaling for thin oxides counters this sensitivity to defects effectively and explains why the percolation model is successful in thin oxides but not in thick oxides. Only defects with the appropriate energy can degrade the breakdown distribution. The required energy is consistent with oxygen vacancy \( E'_v \) defect after capturing a hole and the concentration required is consistent with very high-quality oxide. This explains the consistent low \( \beta \) values for thick oxides universally reported in the literature.

Index Terms: thick oxide; breakdown, oxygen vacancy, percolation model, Weibull slope, trap-assisted-tunneling

Introduction

The need to fight climate change accelerates the push toward the adoption of more efficient, wide bandgap semiconductor based, power electronics to usher in the age of all electric transportations. A common characteristic of these applications is the requirement of high reliability. Gate oxide breakdown is one of the most worrisome failure modes for these devices and the understanding of the breakdown mechanism is of high interest.

Power devices typically use thick gate oxide, from 20 nm to 100 nm. Historically, gate oxide breakdown studies started with thick oxides [1,2]. However, the bulk of understanding evolved as technology nodes advanced, and researchers tended to focus on thinner and thinner oxides. While the theoretical model for thin oxide breakdown is far from established, there are some aspects that enjoy widespread acceptance. One aspect is that breakdown is tightly linked to the current flowing through the oxide during electrical stress. Another is that the percolation model [3,4] is highly successful in explaining time-dependent gate oxide breakdown. It can accurately predict the Weibull shape factor (slope, \( \beta \)) of the thin oxide breakdown distribution. Unfortunately, the model fails to work for thick gate oxides.

The percolation model was inspired by experimental observations from thick oxide breakdown statistics. This model explains breakdown as a defect generation process which is accelerated by electrical stress. The defect generation process proceeds until a critical defect density is established which triggers breakdown [5,6]. Thus, for any given stress field, a critical amount of current, or a critical amount of charge flowing across the oxide is needed to trigger breakdown [7, 8]. Ironically, while the percolation model enjoys great success in thin oxides, thick oxide breakdowns produce \( \beta \) values that are consistently less than the expected value from the percolation model [9-12]. This “failure” of the percolation model led researchers to fall back on the earlier, highly controversial, impact ionization based, model of breakdown for thick oxides [9, 13]. In this work, through simulation of trap-assisted tunneling (TAT), an explanation of the origin of the depressed \( \beta \) value is proposed. The success of this simulation also provides an explanation as to why the same advances in thin oxide growth and consequent reliability improvements do not seem to benefit thick oxides. Moreover, this analysis is consistent with recent literature showing that the “intrinsic” lifetime of SiO\(_2\) grown on SiC is much longer than for oxides grown on silicon [14,15].

Trap-assisted Tunneling (TAT) model

While thermally grown SiO\(_2\) on silicon is famous for being very high quality, even the best films contain a finite number of defects, dictated by thermal dynamics. Careful studies have shown that high-quality thermally grown oxides have oxygen vacancy defects (\( E' \) center precursor) on the order of \( 1 \times 10^{11}/\text{cm}^2 \) [16,17] depending on the thermal history. There are two types of oxygen vacancy point defects, \( E'_v \) and \( E''_v \) at energy \( \sim 4.5 \) eV and \( \sim 8 \) eV respectively below the SiO\(_2\) conduction band edge [18]. Upon capturing a hole, \( E'_v \) would convert to \( E''_v \) - a more stable form of oxygen vacancy defect. Furthermore, the capturing of a hole shifts the energy to \( \sim 2.5 \) eV below the SiO\(_2\) conduction band [18], and renders the defect paramagnetic [17]. Electron injection measurements found that the \( E'_v \) center after hole capture is \( \sim 2.8 \) eV below the oxide conduction band [19], in reasonable agreement with theory. Note that the energy relaxation of the \( E'_v \) center upon capturing a hole will be affected by the film stress which results in a distribution of relaxation energies. As long as the energy is around or above the Fermi level where electron injection originates, the defect will support TAT from the substrate under high electric field. The energy level of \( \sim 2.5 \) eV below the oxide conduction band would allow the same explanation to apply to oxides grown on SiC instead of Silicon. A recent study clearly reports TAT via \( E'_v \) center for SiO\(_2\)/SiC system and the important role of interface strain [20]. This could explain why thick gate oxide breakdown behavior is so similar for the two systems.

It is important to recognize that high electric field stress is known to generate significant hole current flowing across the oxide from anode to cathode, hence under the condition of common breakdown tests the availability of sufficient holes for capture is guaranteed. Combining all the above, the picture is that oxygen vacancies in oxide films will capture holes when the film is under high electrical stress to become TAT active centers, leading to locally increased tunneling current that will shorten the breakdown lifetime.

The linkage of stress-induced-leakage-current (SILC) to breakdown has been investigated extensively [21-25]. SILC is a consequence of TAT and serves as a monitor of defect density increases during electrical stress. While current flowing through the oxide is directly link to oxide...
breakdown, leakage increase by TAT has not been considered as a cause of lifetime shortening. This is likely influenced by the success of the percolation model which assumes defects are created at random locations. A few groups considered defect creation rate is higher near an existing defect in their models [26-30]. They however did not consider how it may affect the value of $\beta$ - an important measure for the validity of the model. Charge-trapping at defect sites leading to the distortion of the oxide field and tunneling current change has been widely accepted [7, 8], but tunneling enhancement through defects as a cause for breakdown distribution change has never been considered.

Since the goal is to examine how defects cause the value of $\beta$ to be degraded in thick oxides, the simulation starts with a constructed oxide breakdown Weibull distribution based on the percolation model (with a characteristic $\beta$ associated with oxide thickness). A defect distribution in the oxide is then introduced and the lifetime shortening due to defects are calculated for the whole population, resulting in a new distribution with reduced characteristic lifetime ($\tau_{63}$) and $\beta$.

Figure 1 illustrates the band diagram associated with the TAT simulation. The blue fuzzy line across the oxide thickness is the defect spatial distribution (assumed uniform) with a Gaussian energy distribution. $\Phi_b$ is the substrate conduction band to SiO$_2$ conduction offset (electron injection barrier height); $\text{Tox}$ is the oxide thickness; $X_T$ is the location of the defect from the interface; $X_T'$ is the tunneling distance of the second step; $V$ is the applied voltage. $\Phi_b - V'$ is the barrier of tunneling from the defect site to the oxide conduction band; $V''$ is the energy distance for the defect from the mean of the defect energy distribution.

The tunneling electron is assumed to be from the Fermi level of the injecting interface. For a given defect density and energy distribution, the probability of finding a defect that has the right energy for TAT can be calculated for each spatial location (distance from the interface). Since defects can only be integers, this probability is realized through randomly distributed defects in a population so that the average of a large numbers of devices equals the probability. TAT is calculated the usual way: the joint probability of the two tunneling steps normalized.

The first step being a direct tunneling situation, the approximate analytical equation of Schuegraf et. al. [31] was employed. Even though there are more accurate models for direct tunneling, they are more complicated and the additional accuracy is not required in this work. The second step is normal Fowler-Nordheim (FN) tunneling. Only elastic TAT is considered because inelastic tunneling will have a much lower probability due to the steep energy relaxation. Note that energy relaxation for inelastic tunneling is often assumed to be shallow, but direct experimental measurement from two different techniques found that the relaxation is ~1.5 eV [32, 33]. From fig. 1 one can see that if deep relaxation occurs, the second step of the TAT will have a much larger barrier. Therefore, for thick oxides, inelastic TAT can be neglected.

For any given stress field, the TAT from defects in the entire triangular region of the oxide barrier are calculated. The largest TAT current is chosen and normalized to the defect-free FN current to obtain the current enhancement factor.

The concept of critical charge to breakdown for a given stress field [3-6] means that current is inversely proportional to breakdown lifetime, current enhancement means defect generation speeds up and lifetime is shortened. One may write,

$$\tau_{63} \propto \frac{1}{I_{\text{tun}}}$$  \hspace{1cm} (1)

$$E = \frac{I_{\text{tAT}}}{I_{\text{FN}}}$$  \hspace{1cm} (2)

Where $\tau_{63}$ is the characteristic time of breakdown in Weibull distribution; $I_{\text{tun}}$ is the current tunneling through oxide under electric field; $I_{\text{tAT}}$ is the TAT current; $I_{\text{FN}}$ is the FN current in the absence of defects and $E$ is the current enhancement factor.

The current enhancement factor cannot be applied directly to (1) to calculate lifetime reduction because TAT current is highly localized to the defect site, which in a previous modeling paper was taken to be 1 nm x 1 nm [34]. The percolation model is based on the idea that the tunneling current will lead to defect generation (Note that the created defects are not necessary the same kind of defects involved in the TAT calculation [35]). This excludes a positive feedback process and that the critical defect density to breakdown is device area-dependent. Thus, to translate the current enhancement factor into lifetime shortening, area effect must be accounted for. A simple way to think of this situation is two parallel capacitors. One is the ultra-small capacitor associated with the defect the other is the rest of the capacitor. While the ultra-small capacitor creates defects faster, it also requires a higher defect density to breakdown. How much higher is determined by area scaling. This area scaling effect is governed by the Weibull statistics.

$$\tau_{63}^{\text{p}} = \tau_{63}^{\text{T}} \left( \frac{A_T}{A_p} \right)^{1/\beta}$$  \hspace{1cm} (3)

Which says the $\tau_{63}^{\text{p}}$ of a capacitor with area ($A_p$) is linked to the $\tau_{63}^{\text{T}}$ of a capacitor with area ($A_T$) through the Weibull slope, $\beta$. We can therefore write:

$$S = \frac{E}{(A_p/A_T)^{\beta}}$$  \hspace{1cm} (4)

Where $S$ is the lifetime shortening factor. $A_p$ is the capacitor area and $A_T$ the defect size. In the simulation, the device area was 1mm$^2$ and the defect area was 1nm$^2$. Thus, the ratio is 1 x 10$^{14}$. One can immediately see that when $\beta$ is large, the effect of the area ratio is negligible. Whereas when $\beta$ is small, the area ratio effectively reduces the impact of the current enhancement factor on lifetime.

**Results and discussion**

Figure 2 shows the impact of defect density on breakdown distribution for a thick oxide (starting $\beta = 40$, ~41 nm thick oxide, left panels) and thin oxide (starting $\beta = 4$, ~5 nm thick oxide, right panels) for the
SiO₂/Si stack. Each distribution consists of 500 devices, the re-sorted simulation result does not vary much from trial to trial. The simulation is repeated for a variety of defect densities consistent for high-quality thermal oxides [16,17]. For the results in fig. 2, the Gaussian defect distribution has mean energy of 0.1 eV (top panels) and 0.2 eV (bottom panels) above the injection Fermi level, and a standard deviation of 0.025 eV. The oxide field is 9 MV/cm. The starting (defect-free) Weibull distribution is label ND with a $t_{63} = 100,000$ s.

Clearly, defects affect the breakdown distribution for thick oxide by both shortening the lifetime and degrading the $\beta$. The degraded distributions are largely parallel with an extracted $\beta$ of ~9, a large reduction from the starting value of 40. Higher mean defect energy (lower panels of fig. 2) degrades the $\beta$ only a little bit more (to ~8) but lifetime reduction is more significant. For the thin oxide, the 0.1 eV mean defect energy has no effect except for the highest defect density case where the lifetime is degraded but not $\beta$. Degradation of lifetime is seen for the higher defect densities when the mean defect energy is 0.2 eV, but the $\beta$ value remains unchanged.

Figure 2. Effect of defect distribution mean energy - Trap-assisted-tunneling effect of breakdown distribution in Weibull plots. The legends are defect densities per cm³ used in the simulation. The defect distribution mean energy is 0.1 eV (upper panels) and 0.2 eV (lower panels) from injection Fermi level. a(1) and a(2) are for thick oxide ($\beta = 40$). b(1) and b(2) are for thin oxide ($\beta = 4$). ND means defect-free.

Figure 3a and 3b show that when the defect energy standard deviation is increased to 0.075 eV the $\beta$ value for thick oxide degrades to ~2.5 while thin oxide has mild lifetime and $\beta$ degradation for the higher defect density cases. Figure 3c and 3d show that reducing oxide field increases the lifetime degradation effect (mean defect energy: 0.1 eV, standard deviation: 0.025 eV, defect density: $1 \times 10^{10}$/cm³). Note that fig 3c and 3d are not showing higher field has longer lifetime. It is showing the degradation from a common starting distribution which is the ND distribution. In reality, the FN tunneling current at 9MV/cm is $\sim 10^{10}$ times higher than at 5MV/cm. Factoring that into account would shift the 5MV/cm plot up by 10 orders of magnitude while the 9MV/cm plot stays where it is.
These simulation results suggest that the low level of defects at or slightly above the Fermi level where tunneling electrons originated has a large effect on the breakdown distribution of thick oxide but would have far less impact on thin oxide. As discussed above, these densities of oxygen vacancy defects are inevitable even for very high-quality oxide. When only these defects exist, the oxide breakdown distribution is commonly classified as “intrinsic”. These results demonstrate how thick oxide “intrinsic” breakdown does not follow the prediction of the percolation model. It also explains why the same process improvements that led the semiconductor industry to routinely grow high quality thin gate oxide do not seem to help the quality of thick gate oxides. This model shows that perhaps the thick oxides have achieved the same quality already.

The simulation results show that defect density in general is not the main factor that affects \( \beta \) for thick oxide, nor is the mean defect energy. Instead, the standard deviation of the Gaussian distribution of defect energy plays a much bigger role. From fig. 1 this can be understood because the peak of the defect distribution (for the given mean energy) is distinct from the peak TAT location under most applied field conditions. Thus, it is the tail of the defect energy distribution that plays the main role and therefore the standard deviation is the most important factor. One might think that increasing the mean energy but decreasing the defect density could produce similar results (and it does). However, the standard deviation will have to be unreasonably small to keep \( \beta \) from going too low. Literature reported \( \beta \) values for thick oxides range from less than 2 to more than 10. This suggests that in some cases the standard deviation may exceed 0.1 eV. For thermally grown oxides, the thermal history controls the defect density [16, 17]. However, the same thermodynamic argument would lead us to expect the mean defect energy and the standard deviation are both controlled by thermal history as well. Intuitively, sufficient time for equilibrium to establish at lower temperatures should lower the defect density, lower the defect energy (relaxes to the bottom of the energy well), and sharpen the energy distribution.

The spatial distribution of the defect has been assumed to be uniform in this simulation. There is reason to believe that a concentration gradient should exist [36] with higher concentrations at the interface. From a stress consideration, the thermal expansion mismatch between SiO\(_2\) and the substrate leaves the greatest residual stress at the interface upon cooling below the “flow” temperature. This stress location would also suggest higher defect densities, higher defect mean energies and higher energy dispersions near the interface. At the same time, only defects near the interface (fig. 1) really matters. While the non-uniform distribution will impact the simulation, the basic conclusion is not expected to change.

These findings may also explain recent reports of thick SiO\(_2\)/SiC devices with intrinsic lifetimes longer than the conventional thick SiO\(_2\)/Si counterparts [14, 15]. These SiO\(_2\)/SiC devices have higher \( \beta \) values for their breakdown distributions (comparing to the thick SiO\(_2\)/Si values reported in the literature). This indicates a higher quality and therefore longer life that is more than compensate the expected shorter life from the mildly smaller electron injection barrier.

If TAT from existing defects shortens lifetime, it begs the question of the TAT effect when more and more defects are created by electrical stress. A feedback runaway condition seems to exist but is not reported experimentally. The fact that only defects at the right energy and right location can support efficient TAT may offer an explanation. There is some evidence that the stress-induced defects and the defects that leads to breakdown are not the same [35]. This, however, cannot rule out small percentage of the stress-induced defects would have the right energy and right location. The buildup of these defects will be slow so any feedback enhancement will be mild and will occur only after longer
stress durations. In addition, it can only happen in thick oxide cases where the existing data cannot rule out this possibility.

Since the explanation for low $\beta$ value is base on oxide defects slightly above the Fermi level of the injecting electrons, it is fair to question whether such phenomena happen to power devices based on materials other than Si or SiC. One cannot answer this question until much more studies both in breakdown distribution as well as defects in the oxide that can be accessed by the inversion charges for the material in question.

### Conclusion

The basis of the percolation model: defect buildup under electrical stress until a critical density is reached leading to breakdown works for thick gate oxides. The inability of the percolation model to predict the $\beta$ of thick oxide breakdown is explained by the hypersensitivity of thick oxide breakdown to low densities of defects that are inevitable even in the best of thermally grown oxide films. Such hypersensitivity also explains why the growth-derived lifetime improvements employed in high quality thin gates oxides have little impact on thick oxides of the same apparent quality. It further explains why, even with a reduced electron tunneling barrier, having larger $\beta$ allows thick oxide on SiC to have intrinsic breakdown lifetime much longer than on silicon.

### References


