Modeling Early Breakdown Failures of Gate Oxide in SiC Power MOSFETs

Zakariae Chbili, *Member, IEEE*, Asahiko Matsuda, Jaafar Chbili, Jason T. Ryan, *Member, IEEE*, Jason P. Campbell, *Member, IEEE*, Mhamed Lahbabi, Dimitris E. Ioannou, *Fellow, IEEE*, and Kin P. Cheung, *Senior Member, IEEE*

***Abstract*— One of the most serious technology roadblocks for SiC DMOSFETs is the significant occurrence of early failures in time-dependent-dielectric-breakdown testing. Conventional screening methods have proved ineffective, because the remaining population is still plagued with poor reliability. The traditional local thinning model for extrinsic (early) failures, which guides the screening through burn-in measures, simply does not work. The fact that improved cleanliness control in the fabrication process does little to reduce early failures also suggests that local thinning due to contamination is not the root cause. In this paper, we propose a new lucky defect model where bulk defects in the gate oxide, introduced during growth, are responsible for the early failures. We argue that a local increase in leakage current via trap-assisted tunneling leads to early oxide breakdown. This argument is supported with oxide breakdown observations in SiC/SiO_2 DMOSFETs, as well as simulations that examine various defect distributions and their impact on the resultant early failure distributions.**

***Index Terms*— Burn-in, DMOSFET, early failures, power MOSFET, reliability, SiC, time-dependent-dielectric-breakdown (TDDB).**

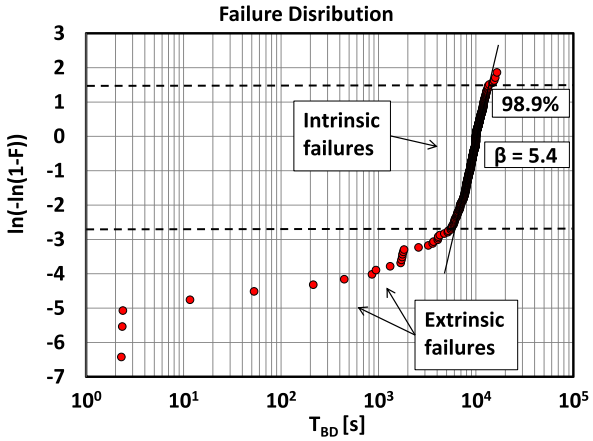

Fig. 1. TDDB results on a total of 432 SiC DMOSFET (200 μm $\times$ 200 μm) with a 50-nm-thick oxide SiO_2/SiC. The collective data from several TDDB test fields (6.8–10 MV cm^{-1}) and temperatures (25 °C–275 °C) were normalized to a breakdown time of 10^4 s. Most of the failures follow an intrinsic lifetime (straight line) with a significant tail of extrinsic failures.

I. Introduction

SILICON carbide (SiC) DMOSFETs are poised to replace conventional silicon devices in power applications due to lower switching losses and higher operational temperatures [1], [2]. However, the quality of the gate oxide and the oxide/SiC interface continue to cause serious reliability concerns that could prevent SiC-based technology from widespread adoption [3]–[6]. There are two major reliability challenges for SiC MOSFETs: 1) bias-temperature instability (BTI) and 2) gate oxide breakdown. Generally speaking, BTI is the result of charge trapping in the oxide and its interface [3], [5]–[7]. Recent advances in annealing and interface passivation techniques [8], [9] have significantly reduced the impact of BTI [2], [10]. Recent studies have also shown that the intrinsic gate oxide breakdown reliability of SiC MOSFETs is similar to thermal oxides grown on silicon [4], [11]–[14] and should be of little concern. However, the extrinsic or early failures greatly determine overall product reliability and have received little or no attention.

While recent time-dependent-dielectric-breakdown (TDDB) studies have shown intrinsic similarities between SiC/SiO_2 and Si/SiO_2 devices and have succeeded in allaying earlier worries of a limiting role for SiC MOSFET oxide breakdown [4], [11]–[14], most of the failure distributions in these studies contain extrinsic breakdowns [4], [13], [14]. In some cases, the early failures were removed to focus on the intrinsic portion of the distribution (fitted to a straight line in the Weibull plot) [4], [14]. When sufficient breakdown events are observed, the failure distribution generally looks like Fig. 1. This is most often characterized by: 1) a relatively tight distribution in which the majority of devices fail

Manuscript received April 11, 2016; revised June 10, 2016; accepted June 27, 2016. Date of publication July 14, 2016; date of current version August 19, 2016. The review of this paper was arranged by Editor G. Meneghesso.

Z. Chbili is with GLOBALFOUNDRIES, Inc., Malta, NY 12020 USA (e-mail: zakariae.chbili@globalfoundries.com).

A. Matsuda is with the National Institute for Material Science, Ibaraki 305-0047, Japan (e-mail: asa.matsuda@gmail.com).

J. Chbili, J. T. Ryan, J. P. Campbell, and K. P. Cheung are with the Semiconductor and Dimensional Metrology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: jaafarchbili@gmail.com; jason.ryan@nist.gov; jason.campbell@nist.gov; kin.cheung@nist.gov).

M. Lahbabi is with the Laboratoire Signaux, Système et Composants, Faculté des Sciences et Techniques, Université Sidi Mohammed Ben Abdallah, Fes 2202, Morocco (e-mail: lahbabi_m@yahoo.fr).

D. E. Ioannou is with George Mason University, Fairfax, VA 22030 USA (e-mail: dioannou@gmu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2016.2586483

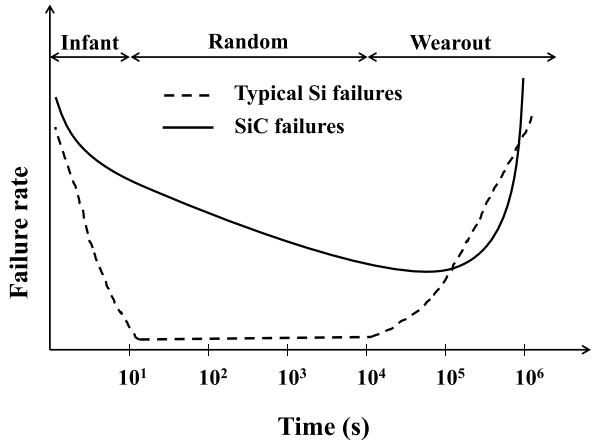

Fig. 2. Illustrative reliability bathtub curve of SiC compared with Si. The high random failure rate in SiC makes burn-in ineffective. No amount of screening can reduce the failure rate low enough to guarantee high reliability.

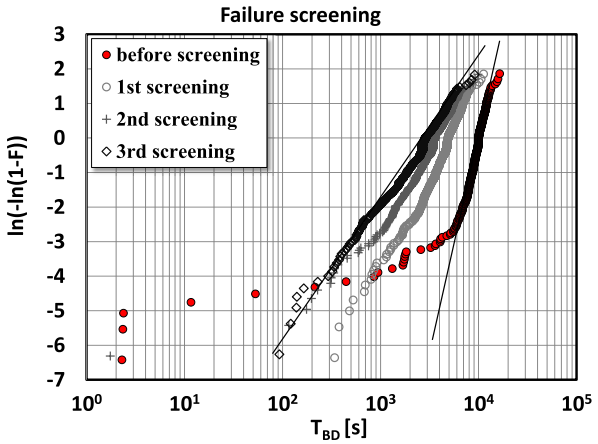

Fig. 3. Multiple screenings of early failures from the failure distribution collected in Fig. 1. Increasing the burn-in time to screen off more early failures (first screening = 5.2 ks, second screening = 6.5 ks, and third screening = 7.3 ks) leaves the remaining lifetime distribution with a shorter time-to-failure and a shallower slope.

(the intrinsic portion) and 2) a long tail of early failures (extrinsic portion) [13]–[15]. Fig. 1 shows the results of TDDB experiments that are explained in detail in Section II, which are representative of those typically observed in SiC TDDB. Early failures are common in product reliability and are usually removed through a burn-in process (screening). The criteria required for successful burn-in is twofold; first, the early failure population must be sufficiently distinct from the intrinsic failure distribution. Second, the intrinsic time-to-breakdown (TBD) must be long enough such that the lifetime consumed by the burn-in is insignificant. The reliability performance that satisfies these criteria is captured by the famous bathtub curve, as shown in Fig. 2 for both typical silicon devices (broken line), and SiC devices (solid curve)—the latter not quite resembling a bathtub.

In Si/SiO_2 MOSFET technology, extrinsic TDDB failures are generally modeled as a variation of localized effective oxide thickness (local thinning) due to contaminants that tend to inhibit oxide growth [16]. At these thinner spots, the result is a much higher effective oxide field and, subsequently, a lifetime much shorter compared with the intrinsic lifetime. Thus, the extrinsic (early) failure population can be removed by screening without significantly altering the intrinsic lifetime. Examining the silicon device curve in Fig. 2, a burn-in time of 10 s will effectively remove all the early failures. Additionally, the 10 s burn-in will have a very small effect on the 10 000 s of lifetime before wear-out begins. Moreover, once wear-out sets in, the failure rate is relatively low, and therefore, high reliability is realized. On the other hand, for the typical SiC/SiO_2 gate oxide failure behavior shown in Fig. 2 (based on the failure distribution shown in Fig. 1), no amount of screening (burn-in) can reduce the failure rate to levels low enough to realize high reliability. Fig. 3 shows this point more clearly by comparing how various levels of screening impacts the failure distribution. Increasing the burn-in time to screen off more early failures (first through third screening) leaves the remaining lifetime distribution with a shorter mean time to failure and a shallower Weibull slope. The shallower Weibull slope is quite alarming and indicates that a lower overall failure rate cannot be achieved by screening in these devices.

The inability to lower the failure rate by screening indicates that for SiC/SiO_2 devices, the early failure distribution is not sufficiently distinct from the intrinsic distribution and suggests a related origin. Using the varying effective local oxide thickness model that is typically invoked to describe these observations would require a continuous thickness variation. Since local thinning is due to localized contamination during oxide growth [16], a continuous thickness distribution is very unlikely. Furthermore, efforts to minimize contamination during oxide growth, as suggested by the oxide thinning approach, have been shown to have little or no effect on the early failure population in SiC [15]. Collectively, these observations strongly suggest that early failures in SiC/SiO_2 gate oxide breakdown are due to some other physical mechanism.

To show that the local thinning model cannot explain the experimental data requires a large number of samples to be tested to achieve good statistics. This was done by Ogier *et al.* [17]. While they found that using a joint probability distribution without assuming any model works well to fit the data, an alternative model was not developed. This lack of an alternative physical model means that there is no guidance of what to do after the reduction of extrinsic failure reaches a limit regardless of how much more effort is put into improving the wafer cleaning. In this paper, we introduce a new defect-based model to explain the occurrence of extrinsic failures even in the absence of contamination. This new model suggests improving the oxide growth process itself as the next step to combat the extrinsic failure problem, once contaminations are thoroughly minimized or eliminated. In other words, the model provides a physics-based path forward to overcoming the extrinsic failures problem.

II. Lucky Defect Model

It is well known that gate oxides grown on SiC have a much higher density of defects than those grown on silicon [18]–[20]. Considering this high defect density reality, we have developed a new lucky defect model to explain the

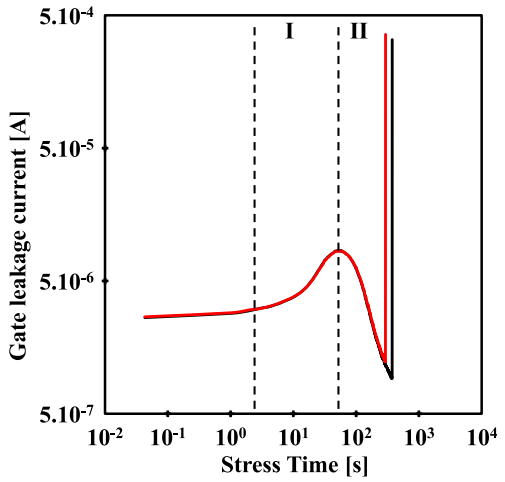

Fig. 4. Measured gate leakage from two identical SiC DMOSFET (200 μm $\times$ 200 μm) with a 50-nm-thick oxide SiO_2/SiC. The current progression clearly shows defect generation during stress and two distinct failure times for two identical devices.

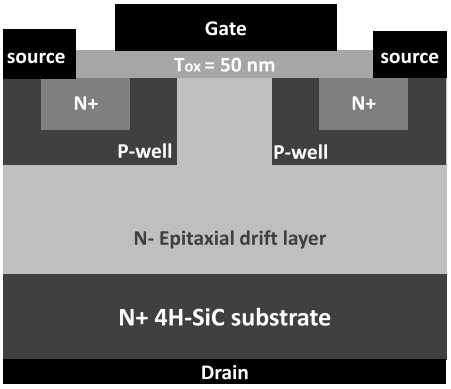

Fig. 5. Device structure of the SiC DMOSFET used in the experiments.

extrinsic (early) failures of SiC/SiO_2 gate oxides. The basic idea of our model is that a lucky defect with the appropriate energy level and spatial location will locally enhance leakage current via trap-assisted tunneling (TAT) during high-field stress. The increased leakage current will lead to an increase in oxide wear-out rate and, therefore, a shorter lifetime. The defect density of the oxide is related to the oxide growth process and, as mentioned earlier, is not dictated by contamination. When contaminants are present, local thinning can, of course, contribute to early failure. However, we assume that the contamination control has been improved [15] to the level that this mode of extrinsic failure is negligible.

The basis for our model is that current flowing through the oxide during stress plays a critical role in TBD. We note that this is not generally accepted for thick oxide (>30 nm) breakdown. The more commonly cited mechanism for thick oxide breakdown is the feedback runaway model triggered by impact ionization [21]–[25]. However, this particular model was challenged since its inception [26]–[29], and the debate has never been completely resolved. The resulting debate compromise is that above a critical field, impact ionization dominates, while below the critical field, breakdown is due to the time-dependent accumulation of stress-induced defects. Though the exact mechanism of defect generation under electrical stress is still under debate, widely accepted experimental evidence [30] suggests that defect generation is related to charge injection. The notion of a critical defect density, and therefore a critical charge to breakdown [30], is the foundation of all later breakdown models and is also the foundation of our model discussed here. Thus, in order to justify the approach of our model, we need to ensure that our TDDB data is collected at fields that support the defect accumulation model.

Fig. 4 shows the progression of measured gate leakage current of two typical 200 μm $\times$ 200 μm SiC/SiO_2 power MOSFETs during TDDB stress (9 MV cm^{-1}, 150 °C, and positive gate bias). The gate current exhibits three different phases. In phase I the gate current increases with stress time, likely due to positive charge (hole) trapping in the oxide. If the breakdown is dominated by impact ionization current, the leakage current will continue to increase until breakdown occurs. However, there is clearly a phase II in which the gate current decreases, likely due to negative charge trapping overtaking hole trapping. Phase III is the abrupt increase in current (breakdown). The increased electron trapping before breakdown indicates that the device undergoes a long period of defect generation and electrons are trapped at these new defect sites. The behavior shown in Fig. 4 is the confirmation we seek in order to justify the approach of our model, as discussed earlier.

A. TDDB Experiments

The devices used in this paper are 200 μm $\times$ 200 μm SiC/SiO_2 DMOSFETs (Fig. 5) with a 50-nm-thick thermally grown SiO_2 gate dielectric. A total of 432 devices were tested and were separated into 24 groups of 18 devices. Each group was tested under a different TDDB stress condition. Positive stress voltage was applied to the gate while the drain was held at ground (the n-layer of the drift region under the gate is at accumulation) and the source was floating. A wide range of oxide fields (6.8–10 MV cm^{-1}) and temperatures (25 °C–275 °C) were studied.

The stress-induced TBD of each group can be plotted in the Weibull scale where the failure distribution is given by

$$F(T_{\mathrm{BD}}) = 1 - \exp\left[-\left(\frac{T_{\mathrm{BD}}}{T_{63\%}}\right)^{\beta}\right] \quad (1)$$

where $F(T_{\mathrm{BD}})$ is the cumulative failure fraction, $T_{63\%}$ is the characteristic time (characteristic time at which 63% of the tested population is broken), and β is the Weibull shape factor (Weibull slope). Assuming the breakdown mechanism is the same for all stress fields and temperatures, we can expect all the groups to have the same β for the intrinsic part of the distribution, since the oxides are identical. On the other hand, the characteristic time ($T_{63\%}$) does change with area, applied field, and temperature. Thus, for the intrinsic part of the distribution, we can normalize the different failure distributions to a given characteristic time, and then plot them together to increase the sample size for better statistics and to highlight the early failure tail.

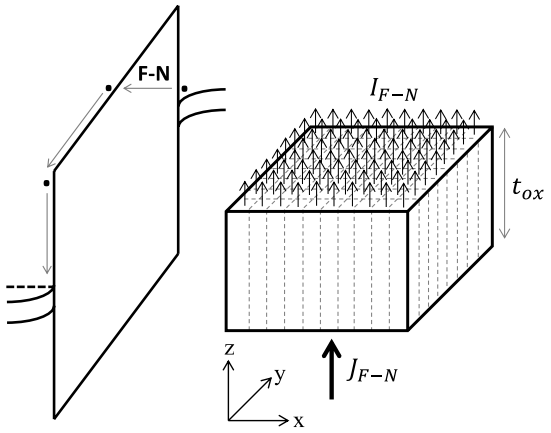

Fig. 6. Uniform F–N tunneling in thick dielectric in the absence of defects.

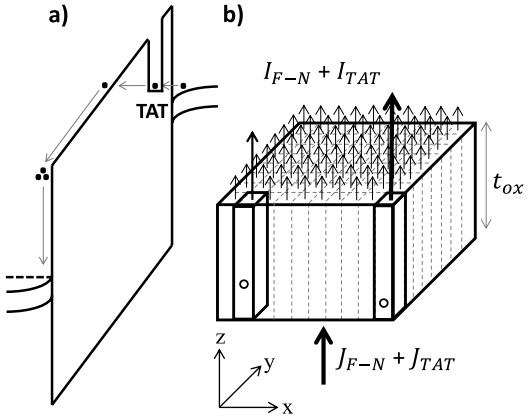

Fig. 7. Tunneling current in the presence of traps. (a) Traps at the lucky energy resulting in (b) local increase in current.

Fig. 1 shows the resulting distribution for all the measured devices normalized to the characteristic lifetime, $T_{63\%} = 10^4$ s. The majority of failures (more than 90% of the population) appear to follow a straight line with a Weibull slope $\beta = 5.4$. This intrinsic failure distribution is, as can be expected, actually affected by the extrinsic early failure distribution. The true intrinsic failure distribution is likely to have a Weibull slope greater than the extracted 5.4 value. To extract the true intrinsic Weibull slope, the data should be analyzed using a joint intrinsic and extrinsic probability analysis [31]. In [4] and [12], the reported β values are much higher, reaching past 20. It is important to note that those results were obtained on capacitors. Our DMOSFET devices underwent many more processing steps and have a much more complicated structure. The reduced β observed here reflects a uniform degradation, but is not the focus of this paper. Here, we focus on the extrinsic or early failure part of the distribution.

B. Model Development

For thick oxides, conduction under high-field stress is accomplished through Fowler–Nordheim (F–N) tunneling. Since we do not assume local thinning, tunneling is uniform across the face of the oxide in the absence of defects, as shown in Fig. 6. When defects are present in the oxide, as is clearly the case for SiO_2 grown on SiC, enhanced tunneling can occur at locations where the defect is luckily located at the appropriate energy and spatial location (distance from the cathode) for optimum TAT efficiency. Since tunneling by definition involves no energy change, the defect energy level must match the electrons in the inversion layer, as shown in Fig. 7(a). Note that this increase in current is only local and has little or no influence on the total gate leakage current measured across the whole area of the device. Note also that the electron does not stay trapped in the defect; instead, it tunnels out to the oxide conduction band. (Otherwise, it is simply charge trapping and not TAT). The important consequences of this are, first, there is no trapped charge to distort the electric field in the oxide, and second, the conduction is no longer uniform across the oxide surface, as shown in Fig. 7(b).

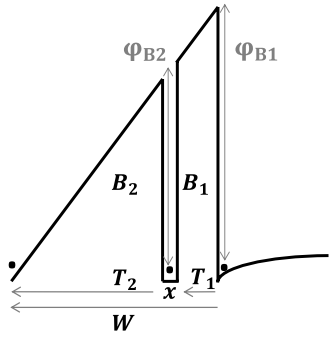

Fig. 8. TAT probability is controlled by the probability to tunnel into the defect (*T1*), and the probability to tunnel out to the oxide conduction band (*T2*). The total tunneling current resulting from TAT is at its maximum when $T1 = T2$.

Since TAT is determined by the joint probability of tunneling into the defect from the inversion layer and then tunneling out from the defect to the conduction band of the oxide, *T1* and *T2*, respectively (Fig. 8), the efficiency is highly dependent on the location of the defect from the cathode. It can be shown that the tunneling current resulting from TAT is at its maximum when the probabilities to tunnel through both barriers are equal ($T1 = T2$), which corresponds to the position x of the defects as

$$x(E_{\text{ox}}) = W(E_{\text{ox}}) * [1 - \sqrt{2}/2] \tag{2}$$

where $W(E_{\text{OX}})$ is the width of the total F–N tunneling barrier at a given field. Note that the location $x(E_{\text{OX}})$ is independent of the barrier height which is a simple consequence of the triangular geometry of the total F–N barrier (T_1 and T_2 are proportional to the barrier areas B_1 and B_2, respectively). Thus, a defect at the lucky location $x(E_{\text{OX}})$, or sweet spot, with the proper energy has the maximum impact on the tunneling current. The lucky energy level and spatial location is, of course, oxide field-dependent.

The lucky energy and location both have distributions that define their values. For energy, the distribution is $2kT/q$ wide centered on the inversion layer energy level, where k is Boltzmann's constant and T is temperature. For location, the effectiveness decays exponentially away from the

sweet spot, but never reaches zero. Both must be accounted for in the subsequent model simulations.

While TAT is modeled as a two steps tunneling process, it is not an inelastic tunneling process. The treatment is in fact for a resonance tunneling process. The charge does not spend any appreciable time at the trap and does not get trapped at the defect site. Consequently, it causes no field distortion.

The best established experimental understanding of oxide breakdown is that there is a critical charge (conducted through the dielectric) to breakdown for a given field and a given oxide [30]–[33]. Thus, we can link increases in conduction, without field distortion, to a faster accumulation of the required critical charge to breakdown. To simplify the model calculations, we assume that the total tunneling current change during stress time, as shown in Fig. 4, remains constant in shape, with an additional component of defect mediated TAT. While defects are indeed created during stress, resulting in the curve shown in Fig. 4, here we are only modeling the effect of as-grown defects. In other words, the higher tunneling current caused by the defect mediated TAT simply shifts the entire current versus time curve up. Alternatively, we are making the Markov Process assumption for defect generation, common to almost all oxide breakdown models.

Since accumulated charge is a product of the current and stress time, one can now directly translate the increase in tunneling current to the decrease in characteristic breakdown time. We can write

$$T_{\mathrm{BD}}^{*} = T_{\mathrm{BD}} \cdot \left(\frac{J_{\mathrm{FN}}}{J_{\mathrm{FN}} + J_{\mathrm{TAT}}} \right) = \frac{T_{\mathrm{BD}}}{\eta} \tag{3}$$

where T_{BD}^{*} is the new reduced characteristic lifetime and η is the current enhancement factor. This lifetime shortening idea is straightforward, but not correct. Here, we have used the characteristic breakdown time T_{BD} of the entire device area and the tunneling current J_{FN} of the entire device area. The highly localized nature of the conduction enhancement must also be incorporated into the description.

Oxide breakdown is commonly described by a weakest link model [34]. In other words, one can think of a single large device with area A as n smaller devices each with area A/n, connected in parallel. If any one of the n smaller device fails, the whole device fails. In general, smaller devices tend to have a longer time to failure. This is true for extrinsic failures caused by contamination as well as intrinsic failures due to the accumulation of charge. While the area effect for extrinsic failures is intuitive, the area effect for intrinsic failures is not. One useful explanation was given by the percolation model for breakdown [35].

In Weibull statistics, this area effect is accounted for by area scaling according to the Weibull slope

$$T_{\mathrm{BD1}} = T_{\mathrm{BD2}} \cdot \left(\frac{A_1}{A_2} \right)^{1/\beta} \tag{4}$$

where T_{BD1} and T_{BD2} are characteristic breakdown times for devices of area A_1 and A_2, respectively, and β is the Weibull slope of the intrinsic breakdown distribution [34]. Thus, T_{BD} in (3) must be adjusted to the size of the local conduction enhancement area. Finally, we can write for a single lucky defect in a device with area A

$$T_{\mathrm{BD}}^{*} = T_{\mathrm{BD}} \cdot \left(\frac{J_{\mathrm{FN}}}{J_{\mathrm{FN}} + J_{\mathrm{TAT}}} \right) \left(\frac{A}{A_{\mathrm{DF}}} \right)^{\frac{1}{\beta}} = T_{\mathrm{BD}} \cdot \frac{1}{\eta} \left(\frac{A}{A_{\mathrm{DF}}} \right)^{\frac{1}{\beta}} \tag{5}$$

where A_{DF} is the area of the enhanced conduction area due to TAT. Defining an appropriate value for A_{DF} is, admittedly, difficult. A defect is often thought of as a point with no dimension, but this cannot be true. Consider a quantum well with infinitely small dimensions, there can be no energy level in it because of the quantum confinement effect. From this consideration alone, one would conclude that a defect must be at least a fraction of a nanometer in dimension at the point where the lowest energy level is located. For our simulation we choose a physically reasonable size of 1 nm^2. Notice that the size of the defect appears in area scaling only. With a β value greater than 5, a small difference in defect size has very little impact on the simulation result.

It is useful to get a sense of the magnitude of this sweet spot driven current enhancement effect. There are two factors affecting the lifetime; one is the area scaling factor and the other is the current enhancement factor. Considering a typical device of area 10^{-3} cm^2, the area ratio becomes 10^{11} (using a localized defect area of 1 nm × 1 nm). Taking this area ratio to the 1/5 power (assuming $\beta = 5$) yields a factor of approximately 150. The current enhancement factor can be calculated, using the Wentzel–Kramers–Brillouin approximation [36] for F–N tunneling

$$J = \mathrm{A}\mathrm{E}_{\mathrm{OX}}^{2} \exp\left(-\frac{B}{E_{\mathrm{OX}}} \right) \tag{6}$$

$$A = \frac{q^3 m_{\mathrm{SiC}}}{16\pi^2 \hbar m_{\mathrm{ox}} \varphi_b}; \quad B = \frac{4(2m_{\mathrm{ox}})^{1/2}}{3q\hbar} \varphi_b^{3/2}$$

where $m_{\mathrm{ox}} = 0.5\ \mathrm{m}_0$ [37], $m_{\mathrm{SiC}} = 0.39\ \mathrm{m}_0$ [38], and $\varphi_{b1} = 2.7$ eV is the barrier height for injection from the SiC conduction band into SiO_2. For TAT, the barrier height φ_{b2} (Fig. 8) can be calculated using geometry and (2)

$$\varphi_{B2} = \varphi_{B1} \cdot \frac{W(E_{\mathrm{OX}}) - x(E_{\mathrm{OX}})}{W(E_{\mathrm{OX}})} \tag{7}$$

As can be seen from Fig. 8, the first step of the TAT process is a direct tunneling process. While there are models for direct tunneling, the calculation is complex. We avoided this by leveraging $T_1 = T_2$ ($B_1 = B_2$) when the defect is located at the sweet spot. Tunneling through B_2 (with barrier φ_{b2}) is an F–N process that is correctly described by (6).

The enhancement factor as a function of applied field for a 50-nm-thick oxide is shown in Fig. 9 (the effect of flat band voltage has been ignored). As can be seen, in the range that is commonly used for TDDB experiments (6–10 MV cm^{-1}), the enhancement factor varies from 10^4 to 10^6—substantially higher than the area scaling factor. The overall result is a substantial reduction in characteristic breakdown time for the small device of the locally enhanced conduction region. By the weakest link model, this dictates that the large area device, as a whole, has a shorter characteristic lifetime.

To sum up the lucky defect model: when a defect in the oxide has the right energy and the right location,

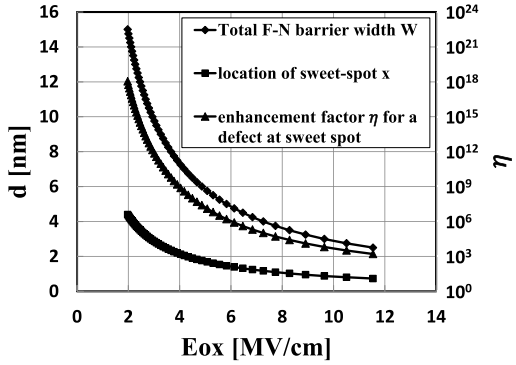

Fig. 9. Change of enhancement factor, total F–N barrier and the location of the sweet-spot all in function of the applied electric field.

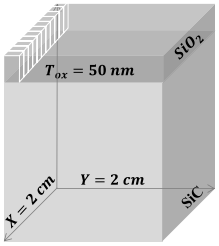

Fig. 10. Simulation structure with a 4-cm^2 wafer featuring 100×100 devices with an area of 200 μm $\times$ 200 μm each.

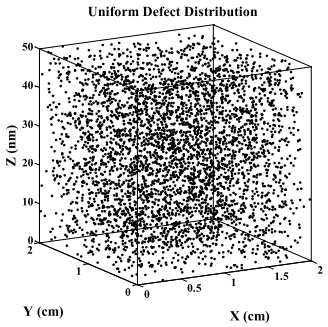

Fig. 11. Simulation structure with a 4-cm^2 wafer featuring 100×100 devices with an area of 200 μm $\times$ 200 μm each.

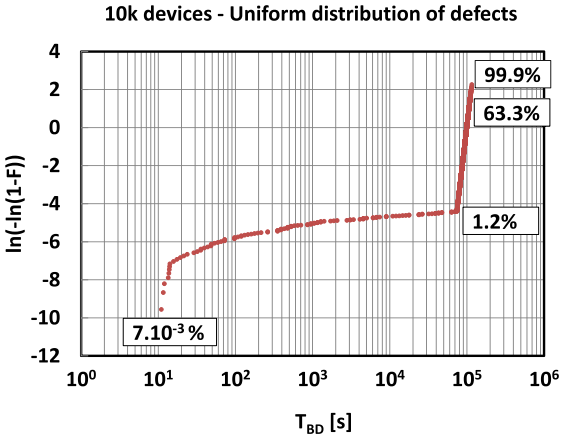

Fig. 12. Simulated failure distribution with a uniform distribution of defects. More than 1% of the simulated devices experience early failures, which results in an early failure tail similar to the experimental results observed in Fig. 1.

local enhancement of tunneling current leads to shorter breakdown time. Depending on the applied field, the lifetime shortening factor can be very large. It is also instructive to examine the impact of the model on thinner oxides. For a thinner oxide, the β value tends to be smaller and therefore the area scaling impact will be larger. At the same time, the TDDB tests are done at higher oxide field for acceleration because the operation field is higher, leading to a smaller current enhancement factor. Thus one expects the lucky defect model will lose its effect for oxides less than 10 nm thick. This is consistent with the large body of reliability literature on thinner oxides.

III. Breakdown Distribution Simulations Using the Lucky Defect Model

A. Methodology

We now examine the effect of the lucky defect model on the gate oxide breakdown distributions in SiC DMOSFETs for the two most often discussed defect distributions. To simplify the simulation, we assume all defects have the appropriate energy to promote tunneling. Furthermore, we include defects that are not quite at the sweet spot, but close enough in location to impact the results. We consider a wafer of area 2 cm $\times$ 2 cm containing 10 000 DMOSFETs, each with an area of 200 μm $\times$ 200 μm (the same size as the measured devices) (Fig. 10).

The first defect profile examined is a uniform distribution throughout the 50-nm-thick SiO_2 dielectric film on a SiC substrate (Fig. 11) with a density of 10^8 cm^{-3}. This is a very low defect density, even for a thermally grown oxide on silicon. However, since only the defects located (energetically) within $2\,kT = 0.104$ eV centered on the top of the inversion layer are effective (and we consider only those that are effective), the implied actual defect density could be orders of magnitude higher. We artificially choose an intrinsic characteristic lifetime of 10^5 s for defect-free devices stressed at $E_{OX} = 8.3$ MV cm^{-1}. The Weibull slope was assumed to be 15. At this field, the width of the total F–N barrier is approximately $W(E_{OX} = 8.3 \text{ MV cm}^{-1}) = 2.5$ nm, and the calculated sweet spot is $x(E_{OX} = 8.3 \text{ MV cm}^{-1}) = 1$ nm from the interface.

To perform the simulation, we construct an intrinsic breakdown time distribution according to the assumed Weibull distribution. We then map the random defect distribution to the device distribution. The result is that some devices will have a lucky defect, while most do not. With a large enough device population and a random defect distribution, this simple approach is statistically valid. For each device, we then identify whether or not there is a defect at or near the sweet spot and calculate the lifetime reduction. In the case that there is more than one defect at or near the sweet spot, we choose only the one closest to the sweet spot for the lifetime calculation. The result is a new breakdown time distribution. Upon sorting the breakdown time, a new Weibull breakdown distribution is created, as shown in Fig. 12.

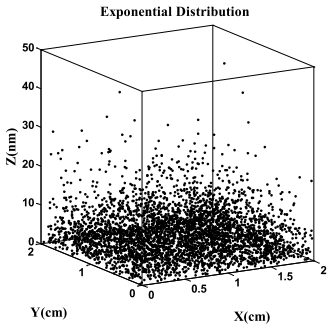

Fig. 13. Random defects with an exponential distribution and a density of defect 10^8 cm^{-3}.

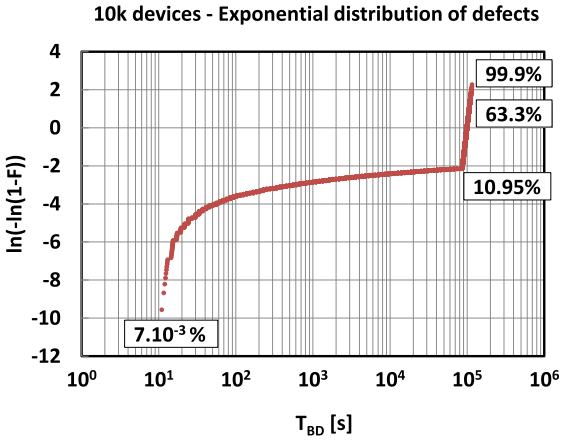

Fig. 14. Simulated failure distribution with an exponential distribution of defects. About 11% of the simulated population experience early failures due to the higher amount of lucky defects in this distribution.

The failure distribution shows an early failure tail of approximately one percent of the population. While not as bad as the distribution of Fig. 1, the basic characteristic is there. One can imagine that a moderate increase in the assumed defect density would easily raise the early failure population closer to those observed experimentally in Fig. 1.

The second defects profile examined is an exponential decay distribution with a density at the SiO_2/SiC interface of 10^8 cm^{-3} and zero at the other interface. Carbon related species are created during processing as an oxidation byproduct. When the oxide is thin, they can easily escape. As the oxide grows thicker, the carbon species cannot escape. This will lead to the kind of exponential distribution examined here. Note that the exponential function is an average across the whole wafer. At any specific location, the defect distribution looks random in thickness, as shown in Fig. 13. The simulation result is shown in Fig. 14 which features a distribution with a much larger percentage of early failures. Compared to the case of uniformly distributed traps, this higher percentage of early failures can be understood by the fact that there are a lot more defects positioned within a small distance from the sweet-spot [$x(E_{OX} = 8.3$ MV cm$^{-1}) = 1$ nm], where they have the best chance to become lucky.

B. Model Implications to TDDB Sample Size and Screening

Most TDDB experiments are performed with a small sample size of 50 devices or less. We now examine the simulated failure distribution while artificially limiting the sample size. Consider a randomly selected group of 50 devices from the 10 000 simulated in Fig. 14. After constructing the Weibull plot, shown in Fig. 15, the distribution shows almost the same percentage of early failures (8%) as the one obtained with a total population of 10 000 devices (11%).

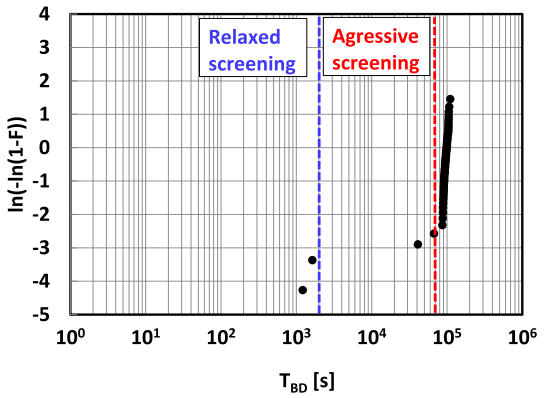

Fig. 15. Failure distribution of a random selection of 50 devices. The results show a limited number of early failures and suggest that a burn-in time of 1.6 ks is a viable option.

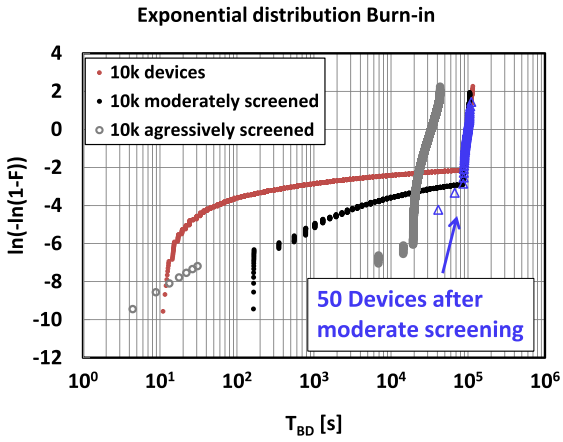

Fig. 16. Effect of 50-devices-based screening on the total population. The burn-in time suggested by the 50-device test does not improve the early failure rate for the remaining of the 10k population.

The limited number of early failures suggests that screening is a viable option to improve reliability. For a relaxed screening (illustrated in Fig. 15), a burn-in time of 1.6 ks seems reasonably tempting and one would expect a post burn-in failure distribution that looks relatively clean, as shown in Fig. 16 (moderately screened 50 devices).

However, the above is a false impression that stems from taking a small sample, and does not reflect the true state of the whole population. When the same 1.6 ks burn-in is applied to the 10 000 simulated devices, the post burn-in distribution is completely different. Fig. 16 shows the effect of the screening with all 10k devices and highlights that burn-in offers no improvement in early failure rate. This echoes the literature results discussed earlier where some reports dropped the early failures from the distribution and concluded that the reliability is high. This is clearly not the case when large

sample populations are used; the ineffectiveness of screening is immediately obvious. This clearly demonstrates that it is important to use a large sample size when investigating oxide reliability. Under sampling can lead to severely flawed conclusions.

Although the presence of lucky defects in SiC devices was shown to result in a large presence of extrinsic defects and lead to false reliability conclusions, the exact nature of these lucky defects in SiC remains unknown. Therefore, it is hard to design an experiment to enhance or decrease the lucky defects in any controlled manner, let alone electrically measuring their contribution. However, proof may come indirectly from process changes that might lead to the elimination of these early failures as was shown in a recent TDDB study [39] on SiC MOS capacitors where an improved process resulted in the elimination of such defects.

IV. Conclusion

In this paper, TAT was presented as a potential culprit of early TDDB failures in silicon carbide DMOSFETs instead of the usual suspect of local thinning due to contamination. Using a new lucky defect model, we demonstrate that the increase of tunneling current due to TAT in oxide bulk defects decreases the time to accumulate the critical charge-to-breakdown and consequently reduces the lifetime of the devices. Through numerical simulations, the effect of two commonly discussed defect distributions were examined and shown to be able to produce the kind of extrinsic failures that are observed and cannot be screened. The success of this model points to a different avenue to solve the early failure problem. Instead of focusing on contamination control, the solution must come from improving the oxide growth process. We would like to note that a parallel can be found in the SiO_2/Si history – in the early 1990s, extrinsic failures were drastically reduced by process improvements.

References

[1] R. Eden. Silicon carbide and gallium nitride power semiconductors. IHS Technology, 2014. [Online]. Available: https://technology.ihs.com/

[2] J. W. Palmour, "Silicon carbide power device development for industrial markets," in *Proc. IEEE Int. IEDM*, Dec. 2014, pp. 1.1.1–1.1.8.

[3] A. J. Lelis *et al.*, "Two-way tunneling model of oxide trap charging and discharging in SiC MOSFETs," *Mater. Sci. Forum*, vols. 717–720, pp. 465–468, 2012.

[4] L. C. Yu, G. T. Dunne, K. S. Matocha, K. P. Cheung, J. S. Suehle, and K. Sheng, "Reliability issues of SiC MOSFETs: A technology for high-temperature environments," *IEEE Trans. Device Mater. Rel.*, vol. 10, no. 4, pp. 418–426, Dec. 2010.

[5] A. J. Lelis *et al.*, "Time dependence of bias-stress-induced SiC MOSFET threshold-voltage instability measurements," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 1835–1840, Aug. 2008.

[6] Z. Chbili *et al.*, "Unusual bias temperature instability in SiC DMOSFET," in *Proc. IEEE Int. Integr. Rel. Workshop Final Rep.*, Oct. 2013, pp. 90–93.

[7] M. J. Marinella *et al.*, "Evidence of negative bias temperature instability in 4H-SiC metal oxide semiconductor capacitors," *Appl. Phys. Lett.*, vol. 91, no. 25, p. 253508, 2007.

[8] G. Y. Chung *et al.*, "Effect of nitric oxide annealing on the interface trap densities near the band edges in the 4H polytype of silicon carbide," *Appl. Phys. Lett.*, vol. 76, no. 13, pp. 1713–1715, 2000.

[9] S. Dhar *et al.*, "Nitrogen and hydrogen induced trap passivation at the SiO_2/4H-SiC interface," *Mater. Sci. Forum*, vols. 527–529, pp. 949–954, 206.

[10] D. R. Hughart, J. D. Flicker, S. DasGupta, S. Atcitty, R. J. Kaplar, and M. J. Marinella, "Progress in SiC MOSFET reliability," *ECS Trans.*, vol. 58, pp. 211–220, Aug. 2013.

[11] L. Yu, K. P. Cheung, J. Campbell, J. S. Suehle, and K. Sheng, "Oxide reliability of SiC MOS devices," in *Proc. IEEE Int. Integr. Rel. Workshop Final Rep.*, Oct. 2013, pp. 141–144.

[12] K. Matocha, G. Dunne, S. Soloviev, and R. Beaupre, "Time-dependent dielectric breakdown of 4H-SiC MOS capacitors and DMOSFETs," *IEEE Trans. Electron Devices*, vol. 55, no. 8, pp. 1830–1834, Aug. 2008.

[13] C. J. Anthony, A. J. Jones, and M. J. Uren, "Quality and reliability of wet and dry oxides on n-type 4H-SiC," *Mat. Sci. Eng.*, vols. 61–62, pp. 460–463, Jul. 1999.

[14] M. Gurfinkel *et al.*, "Time-dependent dielectric breakdown of 4H-SiC/SiO_2 MOS capacitors," *IEEE Trans. Device Mater. Rel.*, vol. 8, no. 4, pp. 635–641, Dec. 2008.

[15] S. Arthur, "GE MOSFET results and issues," presented at the A SiC MOS Program Rev., College Park, MD, USA, 2013.

[16] R. Moazzami and C. Hu, "Projecting gate oxide reliability and optimizing reliability screens," *IEEE Trans. Electron Devices*, vol. 37, no. 7, pp. 1643–1650, Jul. 1990.

[17] J. L. Ogier, R. Degraeve, P. Roussel, G. Groeseneken, and H. E. Maes, "Analysis of the early-failure rate prediction of time-dependent-dielectric breakdown in thin oxides," in *Proc. 25th ESSDERC*, The Hague, The Netherlands, 1995, pp. 299–302.

[18] A. Ohta, K. Makihara, and S. Miyazaki, "Electronic defect states in thermally-grown SiO_2/4H-SiC structure measured by total photoelectron yield spectroscopy," *Microelectron. Eng.*, vol. 147, pp. 264–268, Nov. 2015.

[19] V. V. Afanasev, M. Bassler, G. Pensl, and M. Schulz, "Intrinsic SiC/SiO_2 interface states," *Phys. Status Solidi A*, vol. 162, no. 1, pp. 321–337, 1997.

[20] P. Deák, J. M. Knaup, T. Hornos, C. Thill, A. Gali, and T. Frauenheim, "The mechanism of defect creation and passivation at the SiC/SiO_2 interface," *J. Appl. Phys.*, vol. 40, no. 20, pp. 6242–6253, 2007.

[21] M. Shatzkes and M. Av-Ron, "Impact ionization and positive charge in thin SiO_2 films," *J. Appl. Phys.*, vol. 47, no. 7, pp. 3192–3202, 1976.

[22] P. Solomon and N. Klein, "Impact ionization in silicon dioxide at fields in the breakdown range," *Solid State Commun.*, vol. 17, pp. 1397–1400, Dec. 1975.

[23] B. K. Ridley, "Mechanism of electrical breakdown in SiO_2 films," *J. Appl. Phys.*, vol. 46, no. 3, pp. 998–1007, 1975.

[24] M. Shatzkes, M. Av-Ron, and R. M. Anderson, "On the nature of conduction and switching in SiO_2," *J. Appl. Phys.*, vol. 45, no. 5, pp. 2065–2077, 1974.

[25] T. H. DiStefano and M. Shatzkes, "Impact ionization model for dielectric instability and breakdown," *Appl. Phys. Lett.*, vol. 25, no. 12, pp. 685–687, 1974.

[26] M. V. Fischetti, D. J. DiMaria, S. D. Brorson, T. N. Theis, and J. R. Kirtley, "Theory of high-field electron transport in silicon dioxide," *Phys. Rev. B*, vol. 31, pp. 8124–8142, Jun. 1985.

[27] T. N. Theis, D. J. DiMaria, J. R. Kirtley, and D. W. Dong, "Strong electric field heating of conduction-band electrons in SiO_2," *Phys. Rev. Lett.*, vol. 52, pp. 1445–1448, Apr. 1984.

[28] E. Harari, "Conduction and trapping of electrons in highly stressed ultrathin films of thermal SiO_2," *Appl. Phys. Lett.*, vol. 30, no. 11, pp. 601–603, 1977.

[29] Z. A. Weinberg, "Hole injection and transport in SiO_2 films on Si," *Appl. Phys. Lett.*, vol. 27, no. 8, pp. 437–439, 1975.

[30] Y. Nissan-Cohen, J. Shappir, and D. Frohman-Bentchkowsky, "Trap generation and occupation dynamics in SiO_2 under charge injection stress," *J. Appl. Phys.*, vol. 60, no. 6, pp. 2024–2035, 1986.

[31] R. Degraeve, J. L. Ogier, R. Bellens, P. J. Roussel, G. Groeseneken, and H. E. Maes, "A new model for the field dependence of intrinsic and extrinsic time-dependent dielectric breakdown," *IEEE Trans. Electron Devices*, vol. 45, no. 2, pp. 472–481, Feb. 1998.

[32] R. Degraeve *et al.*, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *IEEE Trans. Electron Devices*, vol. 45, no. 4, pp. 904–911, Apr. 1998.

[33] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, and H. E. Maes, "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides," in *Proc. IEEE Int. IEDM*, Dec. 1995, pp. 863–866.

[34] A. W. Strong *et al.*, *Reliability Wearout Mechanisms in Advanced CMOS Technologies*. New york, NY, USA: Wiley, 2009.

[35] J. Suñé, I. Placencia, N. Barniol, E. Farrés, F. Martín, and X. Aymerich, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films*, vol. 185, pp. 347–362, Mar. 1990.

[36] M. Lenzlinger and E. H. Snow, "Fowler–Nordheim tunneling into thermally grown SiO_2," *J. Appl. Phys.*, vol. 40, no. 1, pp. 278–283, 1969.

[37] R. Ludeke and A. Schenk, "Energy-dependent conduction band mass of SiO_2 determined by ballistic electron emission microscopy," *J. Vac. Sci. Technol. B*, vol. 17, no. 4, pp. 1823–1830, 1999.

[38] S. Potbhare, "Modeling and characterization of 4H-SIC MOSFETs: High field, high temperature, and transient effects," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Maryland, College Park, College Park, MD, USA, 2008.

[39] Z. Chbili *et al.*, "Time dependent dielectric breakdown in high quality SiC MOS capacitors," *Mater. Sci. Forum*, vol. 858, pp. 615–618, 2016.

Authors' photographs and biographies not available at the time of publication.