Explanation for the oxide thickness dependence of breakdown characteristics of metal-oxide-semiconductor structures

D. J. DiMaria and J. H. Stathis
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598
(Received 28 October 1996; accepted for publication 20 March 1997)

Experimental evidence is presented showing that the thickness dependence observed for charge-to-breakdown measurements of very thin oxide layers in metal-oxide-semiconductor structures is correlated with that for the reduction in total generated microscopic defects necessary to induce destruction. These results are related to a percolation model for the formation of paths connecting some of these defects from the cathode to the anode at the time of breakdown. © 1997 American Institute of Physics. [S0003-6951(97)01620-3]

With increasing demand for smaller/faster logic and memory devices, silicon-dioxide (SiO 2 ) layers in the gates of metal-oxide-semiconductor (MOS) structures have been reduced below 10 nm. Thickness down to 2.0 nm (or less) for future logic arrays has also been projected with corresponding reductions in voltage and power. However, reductions in destructive fails with decreasing oxide thickness have not been observed. Typically, many laboratories are generating degradation data like those shown in Fig. 1 where the injected charge to destructive breakdown (Q_{BD}) is shown as a function of gate voltage for oxide thickness in the range of 4.0–6.0 nm. For voltages \( \geq 5 \) V, these data show a pronounced decrease in Q_{BD} with decreasing oxide thickness. This study will focus on an explanation of these observations from measurements of microscopic defect generation caused by hot electrons.

Currently most researchers believe that destructive breakdown of SiO 2 is a two-stage process.\(^1\)\(^-\)\(^8\) In the first stage, microscopic defects (traps and interface states) generated by injected hot electrons are charged up in the oxide layer. During the second stage, a destructive runaway process is triggered "locally" and breakdown occurs. Some think that these local spots are related to microscopic characteristics of the electrode interface with the SiO 2 , such as decorated stacking faults in the Si substrate or grain boundaries in the polycrystalline-silicon (polo-Si) gate.\(^3,4\) However, others have employed statistical arguments for this stage where the formation of local paths through the oxide layer links the contacts together, triggering breakdown.\(^2,8\)

Although there is still a debate concerning the physical nature of the microscopic defect generation, most regard their generation as an important precursor to destructive breakdown.\(^1\)\(^-\)\(^8\) Previously, it had been demonstrated that (Q_{BD}) had a strong correlation to the reciprocal of the "initial" defect generation probability for a wide variation in oxide thickness and/or field.\(^5\) In these simulations, a constant defect density of \( 5 \times 10^{13} \) cm\(^{-2}\) at breakdown was assumed. However for very thin oxides (\( \leq 6 \) nm), this latter assumption could not predict the data in Fig. 1, particularly for voltage magnitudes larger than 5 V. Although some researchers have assumed that there is a critical number of defects at breakdown,\(^1,2,5,7\) only recently have some proposed that this critical number drops significantly on very thin oxides.\(^8\) This study will directly test the validity of this premise, and show that this critical number is approximately independent of electric field, but not oxide thickness (below about 6 nm). Also, statistical considerations will be shown capable of linking an areal generation of microscopic defects with local runaway to break down through paths connecting the defects (like "stepping stones") across the oxide film.

The devices used in these studies were a variety of \( n\)- and \( p\)-channel field-effect-transistors (FETs) with poly-Si gates doped with phosphorous (\( n^+ \)) or boron (\( p^+ \)), respectively, fabricated on the same chip using a CMOS-based technology. These devices will be referred to as \( n\)-FETs/\( n^+ \) and \( p\)-FETs/\( p^+ \) throughout this letter. The thin silicon dioxide (SiO 2 ) layer in these structures was varied between 4.0 and 6.0 nm for a gate area of \( 5 \times 10^{-4} \) cm\(^2\) in a circle FET configuration. For the 5.0 nm gate oxides, some wafers were annealed in \( N_2O \) at 950 °C for 30 min after oxidation. However, the general results reported here were not device specific, and did not depend on gate voltage polarity.

All stressing was performed at high electric fields under Fowler–Nordheim (FN) tunneling conditions. After tunneling into the oxide layer, the near-ballistic electrons were heated by the applied fields to energies nearly equivalent to the potential drop from the entry point to the anode/oxide interface.\(^5\) Total FN-injected electron fluence (Q_{ inj}) was tracked at each gate voltage bias until the structure broke.

*Electronic mail: dimaria@watson.ibm.com

![FIG. 1. Charge-to-breakdown as a function of gate voltage on \( n\)-FETs/\( n^+ \) and \( p\)-FETs/\( p^+ \) for gate oxide thickness varying from 4.0 to 6.0 nm at room temperature.](Image)
down destructively. For sensing trapped oxide charges after FN stress, high-frequency (1 MHz) capacitance–voltage ($C–V$) measurements were performed on the devices after periodically interrupting the FN stress. From the stretch out of these $C-V$ characteristics on the FETs, total interface-state ($N_s$) densities (both fast and slow) were determined. Other quantities related to hot electron induced damage, such as oxide charging or bulk neutral trap generation, were also determined using combinations of $C-V$, current–voltage ($I–V$), and stress-induced-leakage-current (SILC) techniques. These latter studies produced trends similar to those observed for the interface states. To determine initial generation probabilities or the total number of defects at breakdown, buildup of these sites as a function of the hot-electron-fluence was monitored. Generation probabilities were determined from the linear regions of these characteristics at low fluence. The total number of generated microscopic defects at the instant of breakdown was obtained from extrapolation of these characteristics at fluences near destructive failure to $Q_{BD}$.

Examples of the number of generated interface states at breakdown ($N_{sBD}$) as a function of gate voltage are shown in Fig. 2. These data were obtained on some of the same devices as shown in Fig. 1. These data show several trends. First, defect density at breakdown is nearly independent of the applied voltage (or hot electron energy). Second, a strong thickness dependence consistent with that seen in Fig. 1 for charge-to-breakdown is observed above 5 V. However, these data in Fig. 2 do not rule out the possibility that a thickness dependence of the defect generation probability is at least in part responsible for the observations of Fig. 1.

Figure 3 shows the initial defect-generation probability ($P_g$) as a function of gate voltage on the same structures as in Figs. 1 and 2. These data show a strong increase in defect generation above 5 V consistent with the large decrease in $Q_{BD}$ observed in Fig. 1. This experimental relationship has been reported previously for thin oxides (where near-ballistic hot electron transport is occurring) independent of the poly-Si gate doping or voltage polarity. This threshold near 5 V (or 2 eV with respect to the bottom of the conduction band) is believed to be related to hydrogen release from near the anode/oxide interface with subsequent defect growth generated by its motion and/or chemical reaction with existing precursor sites. The data in Fig. 3 do not show a strong relationship to oxide thickness. In this figure, there is no strong thickness dependence between 6 and 8 V as compared to the $Q_{BD}$ data in Fig. 1 over the same voltage range. These data (Fig. 3) alone cannot predict the oxide thickness dependence observed in Fig. 1. However, if values for the reciprocal of the generation probability obtained from Fig. 3 are integrated to the final defect density at breakdown given by the data in Fig. 2 (i.e., $Q_{BD} = qN_{sBD}/P_g$ where $q$ is the magnitude of the charge on an electron), agreement with the data in Fig. 1 can be obtained when the assumption of linearity of $P_g$ up to breakdown is valid.

The data of Fig. 2 indicate that the average $N_{sBD}$ decreases and the scatter in $N_{sBD}$ increases as the oxide thickness decreases. The increased scatter for smaller oxide thickness is a real effect and not experimental error. In the inset to Fig. 2, the same data have been replotted as a cumulative failure (Weibull) plot, where $F(N_{sBD})$ is the fraction of samples for which the interface state density at breakdown is less than $N_{sBD}$. The trends are well explained by a percolation model. Computer simulations have been performed to model breakdown as the formation of a connecting path across an ideal sample comprising a simple-cubic lattice in

![Fig. 2. Total number of interface states (per unit area) at breakdown as a function of gate voltage on the p-FETs/p⁺ used in Fig. 1. Inset: The same data replotted as a cumulative failure (Weibull) plot. The lines are first-order least squares fits.](Image)

![Fig. 3. Interface-state generation probability as a function of gate voltage obtained on the same devices under the same conditions as in Fig. 2.](Image)

![Fig. 4. Predicted and measured defect density at breakdown (63% fails), for sample area $5\times10^{-4}$ cm².](Image)
the shape of a broad, thin sheet. "Defects" are placed randomly by occupying the sites of this lattice with probability \( p \), and breakdown is defined as the formation of a nearest-neighbor-connected network of defects sites across the sheet. The mean defect density at breakdown, \( p_{BD} \), is found to decrease with oxide thickness, as does the slope in the Weibull plot. The thickness dependence of the Weibull slopes was used to normalize the model lattice parameter \( a_0 \), according to which \( a_0 = 1.3 \pm 0.2 \) nm. This physically corresponds to the "size" of a defect, and may be compared to a defect diameter of 0.9 nm found earlier in a similar calculation using randomly placed spheres.\(^8\) With no further adjustable parameters, the result shown in Fig. 4 is obtained for the predicted interface state density at breakdown, scaled to the device area. Here, in order to relate the site occupation fraction \( p \) to the interface state density, \( N_s \) is taken as equal to the defect density in one monolayer of the model lattice, i.e., \( N_s = p/a_0^2 \). The quantitative agreement with experiment is remarkable, especially considering the simplicity of the model. Note that the leveling off of \( N_{s, BD} \) (and hence of \( Q_{BD} \) for constant hot electron energy) for thick oxides (\( >10 \) nm) is correctly predicted as well. The experimental point at large thickness corresponds to the extrapolated interface state density at breakdown which was previously used to successfully model \( Q_{BD} \) in thicker oxides based on linear defect generation rates.\(^5\)

The data presented in Figs. 1–3 imply that fewer total microscopic defects (per unit area) are needed to trigger a breakdown event in very thin oxide layers. This has been explained at least qualitatively from statistical arguments. Destructive paths connecting the anode and cathode become easier to form through the generated defects as the oxide is made thinner.\(^8\) Additionally, below about 10.0 nm, defect distributions (believed to be generated within 5.0 nm of either interface\(^5\)\(^–\)\(^7\)) would overlap and increase the defect density (per unit volume) in the bulk of the oxide. With decreasing film thickness, this increasing volume density of defects would also make the formation of leakage paths between the contacts more probable. Changing oxide processing conditions could also influence the final defect distribution. Evidence for this is shown in Fig. 2 where 5.0-nm-thick standard oxides with and without postoxidation \( \text{N}_2 \text{O} \) annealing are compared. The cumulative-failure distribution for the 5.0 nm \( \text{N}_2 \text{O} \)-annealed oxide has a steeper slope than the standard oxide, although the mean \( N_{s, BD} \) is the same for both. In future percolation calculations, the influence of different spatial and energetic distributions of the microscopic defects will be explored.