Metal-oxide-semiconductor field-effect-transistor substrate current during Fowler–Nordheim tunneling stress and silicon dioxide reliability

Klaus F. Schuegraf(a) and Chenming Hu
Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720

(Received 5 April 1994; accepted for publication 3 June 1994)

The origin of the substrate current of a metal-oxide-semiconductor field-effect transistor when the gate oxide undergoes Fowler–Nordheim stress is investigated. It is also shown that anode hole injection current predicts the breakdown of silicon dioxide between 25 and 130 Å and 2.4 and 12 V. While the measured substrate current is entirely due to anode hole injection for oxides thicker than 55 Å, tunneling by valence-band electrons contributes to the substrate current in thinner oxides. Valence-band electron tunneling current is shown to increase with oxide stressing similar to low-voltage gate oxide leakage; apparently, both are enhanced by trap-assisted tunneling. For oxides of thickness between 25 and 130 Å, the theory of anode hole injection directly verified for oxides thicker than 55 Å is able to model silicon dioxide breakdown accurately.

I. INTRODUCTION

The scaling of device dimensions, resulting in the thinning of oxide dielectrics in VLSI (very large-scale integration) technologies, is driven by the desire to deliver more MOSFET (metal-oxide-semiconductor field-effect-transistor) current and, hence, circuit speed at low voltages, to provide the same capacitance within a smaller area for DRAMs (dynamic random access memories) and to offer lower programming voltages for nonvolatile memories. Due to the high fields present in the silicon oxide mandated by this aggressive dielectric scaling and the presence of oxide defects, catastrophic failure of the dielectric is a major reliability concern. In order to meet future needs of dielectric reliability assurance, a good physical understanding of thin and ultrathin oxide breakdown at low voltages, 3.3 V and below, is required.

Substrate hole current of n-channel MOSFETs has been shown effective in predicting silicon dioxide reliability. This article examines a SiO₂ breakdown model valid between 25 and 130 Å which models this substrate hole current. The relationship between substrate hole current and oxide breakdown is observed for oxides thicker than 55 Å. For thinner oxides, additional contributions to the substrate current due to the tunneling of valence-band electrons obscure the link between generated hole current and oxide breakdown. This article examines these additional contributions to substrate current and shows that prediction of substrate hole current still accurately predicts SiO₂ breakdown despite the role that valence-band electron tunneling plays in limiting the correlation between measurable substrate current and breakdown. First, the anode hole injection model is introduced and shown to predict both the oxide breakdown characteristic and the substrate current for oxides thicker than 55 Å. Next, valence-band electron tunneling is shown to be the source of substrate current for oxides thinner than 45 Å, thereby decoupling the link between substrate current and oxide breakdown. Third, in the transition region between 45 and 55 Å, where the anode hole injection and valence-band electron tunneling currents are approximately of the same magnitude, trap creation during electrical stress can be observed in the substrate current as valence-band-electron trap-assisted tunneling. The stress-created traps greatly enhance the tunneling probability of valence-band electrons whose current begins to exceed the anode hole injection current. This is unimportant for oxide thickness thinner than 45 Å because the fresh valence-band electron tunneling is large to begin with. However, despite these complications, the anode hole injection model can still predict the thickness and voltage dependence of the oxide charge-to-breakdown characteristic Qbd and offers a way to unify many diverse observations into a single model of oxide breakdown.

II. EXPERIMENT

The devices used in this study were either capacitors or transistors with gate oxide thickness varying between 25 and 130 Å. Oxidation process was dry oxidation at 750–950 °C. All devices have an in situ phosphorus-doped polysilicon gate deposited at 605 °C. The capacitors were on n-type (100) 8–12 Ω cm substrate. Constant-voltage stressing of capacitors was performed using a HP4140B picoammeter. Constant-voltage hole separation experiments with transistors were performed using a HP4145B parameter analyzer. All the instruments were controlled by a personal computer using a GPIB (general purpose interface bus) interface.

III. SUBSTRATE CURRENT: ANODE HOLE INJECTION

Recent studies clearly show that hole transport through the oxide precedes breakdown, indicating that holes cause damage to the oxide. The anode hole injection process, depicted in Fig. 1, models the generation of holes for an oxide biased into the high-field tunneling regime, representing the conditions of accelerated testing for determining oxide reliability. A fraction of the tunneling electrons reaching the anode is able to transfer its energy to an electron deep in the valence band. Such an electron is promoted to the conduction-band edge of the anode, thereby creating a “hot” hole, which tunnels back into the oxide. These injected holes...
FIG. 1. Diagram of anode hole injection process. An incident tunneling electron arrives at the anode with energy $E_{\text{gain}}$ to equilibrate. This energy is transferred to a deep valence-band electron, thereby exciting it to the lowest available energy state, the anode conduction band. This excitation creates a hot hole capable of tunneling back into the oxide.

act to increase the current density (at localized spots), probably through hole-induced trap generation, until the final runaway process leads to catastrophic breakdown.

Mathematically, the hole tunneling current is given as

$$J_p = \alpha_p J_n \Theta_p,$$

where $J_n$ is the incident electron tunneling current, $\alpha_p \Theta_p$ is the probability for a hole to be generated and to tunnel through the barrier. The quantum efficiency of the hole generation process is

$$\eta = \alpha_p \Theta_p \exp \left( -\frac{\hbar}{E_{\text{ox}}} \left( \Phi_p (V_{\text{ox}}) \right)^{3/2} \right),$$

(1)

where

$$\hbar = 8 \pi \sqrt{3 m_p \omega_{\text{ox}} / 3 h \nu},$$

$$m_{p,\text{ox}} = 0.2 m_0,$$

and

$$q \Phi_p = E_{\text{gain}} - q \Phi_0 - q \Phi_b - E_{\text{gain}}.$$

The energy gained from the oxide field before arrival at the anode in the Fowler–Nordheim tunneling regime, where $V_{\text{ox}} > \Phi_b / q$, calculated by means of a phenomenological energy relaxation model, is given as

$$\frac{E_{\text{gain}}}{q} = 4 \nu_{\text{ox}} \Xi,$$

(4)

where $\Xi = 15 \text{ Å}$ is the mean-free-electron scattering length in the oxide conduction band. In contrast, in direct tunneling, i.e., for $V_{\text{ox}} < \Phi_b / q$, electrons do not experience such scattering, thus the arrival energy of electrons at the anode is simply

$$E_{\text{gain}} = q V_{\text{ox}} \Phi_b / q.$$

(3)

FIG. 2. Shows the bias configuration of n-MOSFET which enables separate measurement of tunneling electron and injected hole currents. Time-integrated electron tunneling is $Q_{\text{bd}}$, whereas time-integrated hole current is $Q_p$.

Using n-channel MOS transistors biased as in Fig. 2, the tunneling electron flow from the inversion layer to the gate can be easily measured as gate-to-source/drain current, approximately equal to the gate current, and the anode-injected hole current can be measured as the substrate current. Figure 3(a) shows that the tunnel current density $J_g$ decreases by over an order of magnitude during the oxide lifetime due to electron trapping. However, as clearly illustrated in Fig. 3(b), the hole-generation quantum efficiency remains constant for...
Anode Hole Injection Theory

The duration of the oxide lifetime showing that the hole generation rate is strictly determined by the applied bias, consistent with Eq. (1) with \( \alpha_p \approx 0.08 \).

The injected hole fluence increases with time as

\[
Q_p(t) = \int J_p(t) dt
\]

until a critical hole fluence \( Q_p \) is reached, marking the breakdown event. The charge to breakdown \( Q_{bd} \) follows as

\[
Q_{bd} = Q_p \frac{1}{\alpha_p} \exp \left( \frac{\beta}{F_{ox}} [\Phi_p(V_{ox})]^{3/2} \right). \tag{5}
\]

Figure 4 not only demonstrates the proportionality between \( Q_{bd} \) and the inverse of the hole injection quantum efficiency, i.e., \( 1/(J_p/J_n) \), but also demonstrates the effectiveness of the anode hole injection theory postulated by Eqs. (1) and (5). Figure 5 shows that the critical hole fluence at breakdown, \( Q_p \approx 0.1 \text{ C/cm}^2 \), is independent of the stress voltage, while \( Q_{bd} \) decreases with increasing stress voltage according to Eq. (5) for oxides with thickness between 55 and 100 Å. For thinner oxides, the appearance of valence-band-electron tunneling currents or valence-band-electron trap-assisted tunneling currents must be considered as done in the following section; however, subsequent sections show that the equations of the anode hole injection model retain their ability to predict thin oxide reliability parameters.

IV. SUBSTRATE CURRENT: VALENCE-BAND-ELECTRON TUNNELING

For oxides thinner than 45 Å, the substrate current measured by the hole separation technique results from tunneling by valence-band electrons as shown in Fig. 6. Figure 6 illustrates a situation where electrons may tunnel both from the cathode conduction and valence bands. In the hole separation bias configuration, conduction-band electrons are supplied by the source and drain. Tunneling valence-band electrons leave behind holes which are driven from the surface by the depletion-region field and collected as substrate current. Valence-band-electron tunneling is only detectable in such very thin oxides because the tunneling probability of valence-band electrons with barrier height of 4.25 eV in thin oxides is sufficiently high to dominate the anode hole injec-

FIG. 4. Shows that both \( Q_{bd} \) and the inverse of the hole injection quantum efficiency, \( 1/(J_p/J_n) \), follow the same bias dependence.

FIG. 5. Shows that while the critical hole fluence at breakdown \( Q_p \) is almost independent of the stress voltage, the charge to breakdown \( Q_{bd} \) decreases with increasing stress voltage according to the anode hole injection model.

FIG. 6. Band diagram illustrating the situation where electrons may tunnel from both the cathode conduction and valence bands. The valence-band component is directly measurable as \( n \)-MOSFET substrate current for oxides thinner than 45 Å.

FIG. 7. Valence-band-electron tunneling is shown to be much larger than the amount of anode hole injection current predicted from theory in a 32 Å sample.
Anode Hole Injection Model

\[ X_{ox} = 3.2 \text{ nm} \]

\[ \text{Charge to Breakdown (Coulombs/cm}^2) \]

\[ Q_{bd} \]

\[ Q_s \]

\[ Q_p \]

FIG. 8. Charge to breakdown \( Q_{bd} \) follows the prediction of the anode hole injection model although the time-integrated substrate current at breakdown \( Q_s \) is no longer constant because the substrate current reflects valence-band-electron tunneling rather than anode hole injection. At higher stress voltage, \( Q_s \) approaches the \( Q_p \) value needed as a fitting parameter for the \( Q_{bd} \) characteristic because the valence-band-electron tunneling and anode hole injection currents become equal.

FIG. 9. Comparison of experimentally measured \( Q_p \) values using hole separation and those deduced from fitting capacitor breakdown data.

V. SUBSTRATE CURRENT: VALENCE-BAND-ELECTRON TRAP-ASSISTED TUNNELING

For transistors with oxide thickness between 45 and 55 Å, the substrate current measurement allows direct observation of trap creation during high-field stress. Trap-assisted tunneling of valence-band electrons arises, dominating the anode hole injection substrate current [which was shown to remain constant with time in Fig. 3(b)] measured from a fresh device. Figure 10 displays the time evolution of the quantum hole generation efficiency as in Fig. 3(b) for a 49 Å device. For higher bias, the ratio \( J_p/J_n \) is time invariant, following the anode hole injection model. For lower fields, the substrate current increases during stress due to an additional current contribution, the tunneling of valence-band electrons through stress-induced traps. Figure 11 compares electron and substrate currents of a device before and after Fowler–Nordheim stress where a stress-induced leakage component is seen not only in the electron tunneling current, but also in the substrate current. The stress-induced leakage "tail" in the electron tunneling characteristic results from trap-assisted tunneling where electrons direct tunneling into a band of traps near the cathode observe a Fowler–Nordheim-type current relationship with an effectively reduced barrier height. The trap-assisted tunneling of valence-band electrons explains the increase of hole injection quantum efficiency in Fig. 10. Only the lower-bias levels exhibit this increase since the anode hole injection component dominates the stress-induced leakage component at higher biases as shown in Fig. 11.

FIG. 10. Illustrates the time evolution of \( I_p/I_n \) for a 49 Å oxide. At higher bias voltages, this ratio remains time invariant as in thicker oxides. At lower voltages, the \( I_p \) increases due to the additional component of valence-band tunneling through stress-induced traps.
Tunneling Electron Current, \( J_e \), Stress Induced Leakage "Tail"

Valence Band Tunneling

2 4 6

Oxide Voltage (Volts)

FIG. 11. Illustrates the appearance of stress-induced leakage tails in both the tunneling electron and substrate current characteristics. The tail in the conduction-band-tunneling electron characteristic is due to trap-assisted tunneling. The tail in the substrate current characteristic is due to trap-assisted valence-band-electron tunneling.

Again, Eq. (5) accurately predicts the \( Q_{bd} \) characteristic as shown in Fig. 12. Since the time-integrated substrate current \( Q_s \) measures the substrate current contributions of both the anode hole injection and the valence-band-electron trap-assisted-tunneling processes, it is not constant for all biases. However, using the high-field \( Q_s \) of Fig. 12 \((0.01 \text{ C/cm}^2)\) in Eq. (5) enables modeling the \( Q_{bd} \) characteristic in Fig. 12.

VI. OXIDE RELIABILITY PREDICTION: ANODE HOLE INJECTION MODEL

Figure 13 summarizes the power of the anode hole injection model to predict the bias dependence of thin oxide

\[ Q_{bd} \]

for oxides between 25 and 100 Å. The rapidly rising \( Q_{bd} \) behavior in thinner oxides \((Q_{bd} \text{ exceeds } 10^7 \text{ C/cm}^2 \text{ for } 25 \text{ Å oxide at } 2.4 \text{ V})\) can be attributed to the fact that the hot hole energy \( E_{gain} \) becomes quite small and more sensitive to \( V_ox \) when scattering becomes weaker. A small hot hole energy means many electrons may flow through the oxide (large \( Q_{bd} \)) before the critical hole fluence of anode-injected holes \( Q_p \) is reached. Time to breakdown can also be modeled.²

VII. DISCUSSION AND CONCLUSIONS

While the preceding discussion focused on the substrate hole current and its role limiting the direct verification of the quantitative model of thin oxide breakdown characteristics, it only hints at the physical mechanism responsible for SiO₂ breakdown. During the course of electrical stress, oxide damage manifests itself as generated interface traps \( N_{It} \), generated bulk traps,¹⁰⁻¹³,¹² three level stress-induced leakage,¹⁷,¹⁹ and finally, catastrophic breakdown.

Many studies have correlated a particular observation of these manifestations to oxide breakdown and drew conclusions about the cause of oxide wearout. For instance, Harari noted a fixed amount of electron trapping before breakdown and reasoned that electron trapping establishes high internal fields which lead to material rupture. This model fails to account for the role of holes in the breakdown process. Hot hole injection by non-Fowler–Nordheim processes such as junction avalanche show that hole fluence, not electron fluence, is correlated to oxide breakdown. Hole fluence is also much more deleterious to oxide integrity than electron fluence in that the oxide can sustain a fluence of about 10⁰ times more electrons than holes. (This agrees with Fig. 4 where one hole is injected into the oxide for every several hundred electrons tunneling through the oxide.) Others have suggested that current density can be enhanced by hole trapping and this process leads to oxide wearout. This scenario fails to account for the fact that trapped holes are converted into electrons traps and that hole trapping eventually leads to net negative charge trapping in the oxide.

Others have shown the correlation between interface trap generation rate and breakdown under static Fowler–
Nordheim stress,\textsuperscript{22,28} that is, the higher the $N_{\text{eff}}$ generation rate, the lower the charge to breakdown. Yet, dynamic bipolar stress generates considerably more interface states than static stress, while also yielding a longer, not shorter, breakdown lifetime.\textsuperscript{29,30} Moreover, Chen et al.\textsuperscript{4} showed that large interface-state generation under channel hot-carrier stress did not degrade oxide breakdown integrity. These experiments serve to argue against a causal relationship between interface traps and breakdown.

Breakdown appears to be a localized process triggered by a local increase in stress current density. Although the small area of this locally large current density precludes direct measurement of current enhancement, the study of low-level stress-induced leakage\textsuperscript{17-19} seems to indicate a conduction mechanism of trap-assisted tunneling through neutral traps. Such neutral traps\textsuperscript{31} created by a two-step process of hole trapping and subsequent recombination with electrons\textsuperscript{11-13} may serve to explain the localized increase in current density, be it through the mechanisms of charge-assisted tunneling,\textsuperscript{25} trap-assisted tunneling,\textsuperscript{20} or a closely related resonant tunneling.\textsuperscript{31} Since hole fluence leads to the generation of bulk and interface electron traps,\textsuperscript{10-13} it is not surprising that there is often, although not always, a correlation between oxide breakdown and electron trapping or interface-trap generation. In other words, holes in the oxide are the agent responsible for electron trap and interface-trap generation as well as breakdown. There, a casual observation might conclude that breakdown is correlated with any of these other quantities. Recent models emphasizing the motion of hydrogen in trap generation and breakdown can also be linked to hole injection.\textsuperscript{22,32,33} Another model proposes that bond breaking by the energetic tunneling electrons forms a conductive path from anode to cathode.\textsuperscript{34} This may be describing the same hole-induced damage and leakage model espoused here.

Others believe that heating is the final cause of oxide breakdown. A postbreakdown failure analysis would reveal that breakdown is a point of localized melting,\textsuperscript{35} reflecting a large energy discharge at breakdown. However, since the discharge process is virtually instantaneous, compared with the damage accumulation time, the oxide breakdown lifetime is determined by the accumulation of the critical damaging hole fluence. Discharge is but the firework display marking the end of the oxide lifetime.

The anode hole injection model includes and integrates all these diverse and valid observations about the breakdown process. Breakdown is a two-step process. The first part is a long process, possibly spanning years, where the oxide is slowly damaged under electrical stress. The second step is a very short runaway event, on the order of perhaps a $\mu$s, where a rapid final acceleration of damage due to electrical and/or thermal runaway leads to the formation of a permanent conductive path through the oxide. The injection of holes into the oxide is a precursor to oxide breakdown. Hole injection and trapping lead to the generation of both bulk electron and interface traps and possibly the release of hydrogen.\textsuperscript{10-13,27,33} Therefore, hole injection, electron trapping, interface trapping, and perhaps even hydrogen motion can all be correlated to oxide wearout unless special situations such as dynamic stress and hot hole injection are considered. The electron traps, so produced, can enhance current density through the oxide. The enhanced current density generates more damage. This positive feedback process causes destructive breakdown to occur at a weak spot.

**ACKNOWLEDGMENTS**

This research is supported by a JSEP Fellowship under ARO Contract No. DAAL-03-92-G-0242, AMD, Hewlett-Packard, Philips, Rockwell, and Texas Instruments under MICRO, ISTO/SDIO administered by QNR under Contract No. N00014-92 J 1757, and SRC under Contract No. 93-MJ-148.