

CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics

by E. Y. Wu
E. J. Nowak
A. Vayshenker
W. L. Lai
D. L. Harmon

The limitations of reliability of silicon dioxide dielectric for future CMOS scaling are investigated. Several critical aspects are examined, and new experimental results are used to form an empirical approach to a theoretical framework upon which the data is interpreted. Experimental data over a wide range of oxide thickness (T_{ox}), voltage, and temperature were gathered using structures with a wide range of gate-oxide areas, and over very long stress times. This work resolves seemingly contradictory observations regarding the temperature dependence of oxide breakdown. On the basis of these results, a unified, global picture of oxide breakdown is constructed, and the resulting model is applied to project reliability limits for the wear-out of silicon dioxide. It is concluded that silicon-dioxide-based materials can provide a reliable gate dielectric, even to a thickness of 1 nm, and that CMOS scaling may well be viable to the 50-nm-technology node using silicon-dioxide-based gate insulators.

1. Introduction

Silicon dioxide has served for more than three decades as the gate insulator responsible for blocking current in insulated gate field-effect transistor (IGFET) channels from the gate electrode in CMOS devices. The reason for the nearly exclusive use of silicon dioxide in this application is that silicon dioxide uniquely possesses the required combination of several properties: good mobility of holes and electrons flowing in silicon at the silicon dioxide interface, ability to keep electronic states (surface states) at this interface low, relatively low trapping rates of holes and electrons, and excellent compatibility with CMOS processing. Since the gate insulator also serves to couple the electric potential from the gate electrode to the channel of the IGFET, the best control of the channel by the gate is obtained when the gate insulator is made as thin as possible.

A lower bound for the useful gate insulator thickness is presented by the empirical observation that when a voltage is maintained on a large number of gates (which are initially excellent electrical insulators), after some time, some fraction of these gates begin to conduct current from the channel (or source or drain) of the FET

©Copyright 2002 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/02/\$5.00 © 2002 IBM

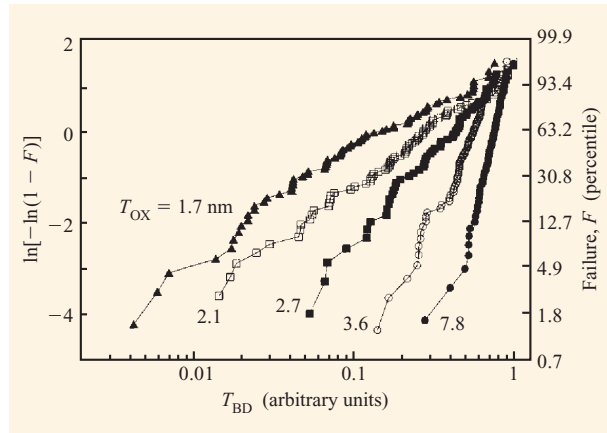


Figure 1

Typical normalized T_{BD} distributions for oxide thickness from 1.7 nm to 7.8 nm. The label on the right vertical axis is cumulative failure probability F as given in Equation (1), in percentiles. The label on the left side is expressed as $\ln[-\ln(1-F)]$, the so-called Weibull scale. Reprinted with permission from [11].

as a result of gate-oxide breakdown. This onset of gate conduction can result in a failure of the circuit employing the failing FET. This behavior of oxide breakdown is usually treated statistically by using a Weibull distribution, given by

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

where F is the cumulative failure probability and T_{BD} is the time-to-breakdown. The characteristic time, α , is the time-to-breakdown at approximately the 63rd percentile, and β is the Weibull shape factor, often called Weibull slope. Typical T_{BD} (or t_{BD}) distributions are given in **Figure 1** for several oxide thicknesses. Because of the weakest-link character of oxide breakdown, the area scaling of T_{BD} is given by

$$T_{BD}(A_1) = T_{BD}(A_2) \left[\frac{A_2}{A_1} \right]^{1/\beta}, \quad (2)$$

where A_1 and A_2 correspond to two different areas of oxide films. Equations (1) and (2) apply equally well when charge-to-breakdown (Q_{BD}) is chosen as the breakdown variable (in place of T_{BD}).

CMOS processes in current production feature oxides as thin as 1.7 nm, and it is generally believed that this thickness must be further reduced to enable continued scaling improvements in density, performance, and power in very-large-scale integrated (VLSI) circuits; thus, understanding of the reliability of such thin oxides is critical.

Numerous researchers have projected over the last decade [1–5] that the properties of silicon dioxide gate

dielectrics would present a limit to the scaling of CMOS technology at thicknesses ranging from at least 10 nm to, more recently, 2 nm. While these barriers have been circumvented, tunnel current limitations and newer dielectric reliability projections of the gate insulator continue to threaten an end to CMOS scaling unless a suitable replacement for silicon dioxide is soon developed. Physics-based models [3–5] have been extrapolated to project that the *International Technology Roadmap for Semiconductors* (ITRS) [6] will require either new materials by the 130-nm node or significant modification (of the roadmap) to the thicknesses of silicon-dioxide-based gate dielectrics or to the operating voltages of these technologies. Naturally, these predictions have stimulated vigorous research in this field, with attention directed, in particular, toward new, high- k gate dielectrics.

A different point of view on the question of extendibility of silicon-dioxide-based gate insulators for CMOS scaling beyond the 100-nm-technology node is presented in this paper. In particular, a very simple empirical model for the intrinsic reliability of dielectric is described and used to project maximum reliable operating voltages for SiO_2 -based dielectrics with thicknesses of less than 1 nm. We find that the ITRS roadmap can be realized to at least the 50-nm node using silicon-dioxide-based gate dielectrics, provided that the issues surrounding gate-tunneling current can be satisfactorily resolved by some combination of process and design innovation.

While significant progress has been made because of the large research effort in the area of ultrathin oxide reliability [3–5, 7–35], crucial questions on this subject remain unresolved. For example, despite intensive research spanning more than three decades, no consensus has been reached about the exact oxide-breakdown mechanisms. Several physical breakdown models have been advanced [7–9], but all of these models present significant limitations in providing an adequate explanation of the breakdown data. Two approaches have commonly been employed: 1) theoretical models based on the assumption that some specific species, typically protons or holes, are directly involved in the breakdown process, and 2) models constructed using indirect methods, such as those based on physical interpretations of stress-induced leakage current (SILC) measurements. Quantitative agreement of these approaches with experimental data has not yet been demonstrated.

To construct an empirical picture of oxide breakdown, an experimental approach spanning a period of five years and many thousands of observations was adopted in order to collect breakdown data with great statistical accuracy. To cover the entire range of stress parameters, two

methods were employed; one makes use of structures with differing areas, and the other, long-term stresses. Additionally, focus was placed on the self-consistency of this empirical model, ensuring that the experimental data stands on its own, without reliance on any underlying assumptions about the mechanisms involved. This approach enables the use of this extremely large database to test hypotheses regarding the underlying assumptions and predictions of various physical models and to provide an accurate extrapolation of gate-oxide reliability for sub-1-nm oxides for use in future CMOS technologies.

2. Experiment

A variety of capacitors and standard n-FETs and p-FETs were fabricated using standard IBM CMOS technologies with T_{OX} ranging from 1.2 to 5.0 nm. Constant-voltage stress (CVS) was used, with the initial breakdown event defining the time of oxide breakdown (irrespective of whether the event is a soft or hard breakdown). For thinner oxides, small-area structures were used in order to carefully avoid series resistance effects. Every reported T_{BD} (Q_{BD}) data point was obtained from a distribution with sample sizes between approximately forty samples and as many as hundreds of samples, depending on oxide thickness. This ensures the accuracy of the measurements [10, 11], so that the error bars associated with each reported data point are comparable to the size of symbols plotted in the figures. Throughout this work, whenever necessary, correction to the T_{BD} results for the effects of oxide thickness variation [11] was made using Q_{BD} and the average of initial tunneling currents, $\langle J_0 \rangle$, that is, $T_{\text{BD}} = Q_{\text{BD}} / \langle J_0 \rangle$. The strong area dependence of time-to-breakdown, due to shallow Weibull slopes for ultrathin oxides, was employed to expand the experimental time window by using capacitors with a wide range of areas [19, 20, 26]. This methodology avoids the potential pitfalls associated with interpreting experimental results collected using different stress parameters such as T_{OX} , voltages, and temperature.

3. Thickness dependence of Weibull slopes and time (charge)-to-breakdown

It was recently found that the Weibull slopes of T_{BD} distributions decrease with decreasing oxide thickness [12], as shown in Figure 1. This reduction in Weibull slopes can contribute to significant errors in area scaling and to extrapolation of the stress data to low failure percentiles [4, 13]. Percolation models [12, 13] and a simple analytical model [14] hypothesize that oxide breakdown is triggered by the formation of a conducting path between the cathode and anode. Such a conducting path forms when a critical defect density, N_{BD} , as originally proposed in [15],

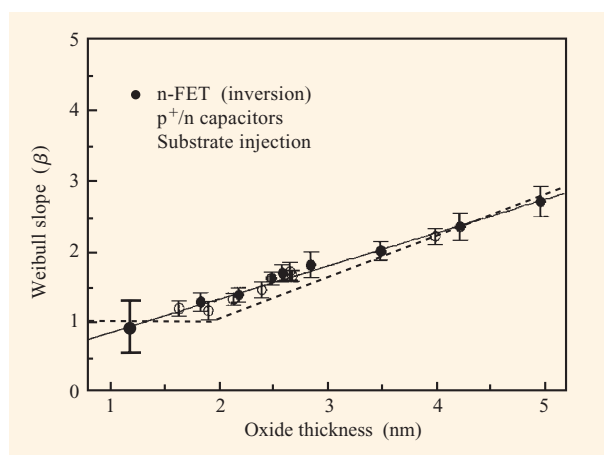


Figure 2

Weibull shape factors (slopes) as a function of oxide thickness. Reprinted with permission from [11].

is reached as a result of defect generation during oxide stress. The reduction of Weibull slopes is attributed to a decrease of N_{BD} [12–14].

Highly accurate determination of Weibull slope directly from T_{BD} distributions requires thousands of samples [10]. To circumvent the enormous effort required to obtain such extreme sample sizes, the area-scaling relation of Equation (2) was employed [11]. The main advantage of this method is that it is much less sensitive to statistical uncertainty and thickness variations, with an achievable accuracy of 5% provided that a sufficiently large area ratio is used [11]. **Figure 2** shows the measured Weibull slopes obtained via this method vs. T_{OX} ranging from 1.25 nm to 4.96 nm at a temperature of 140°C. The dashed curve represents the prediction of a percolation model [13], whereas the solid curve is a linear fit of a recently published analytical cell-based model [14]. General agreement among the experimental data and the two models, as seen in Figure 2, demonstrates the success of this simple geometrical model [12–14].

The decrease of N_{BD} with decreasing T_{OX} also predicts a reduction of Q_{BD} with decreasing T_{OX} , as previously reported [16]. Some typical results are compared with a cell-based model [14] in **Figure 3**. In order to compare our data with the analytical model [14], the independent variable was changed to $1/(T_{\text{OX}} + T_{\text{int}})$. The parameter T_{int} is introduced in the model and interpreted as an interfacial layer thickness [14]. Good agreement of this model with the experimental data supports the general approach of the model [14]. The thickness scaling of T_{BD} is similar to that found for Q_{BD} , but in this case the thickness dependence of the tunnel current $J(T_{\text{OX}})$ must also be considered.

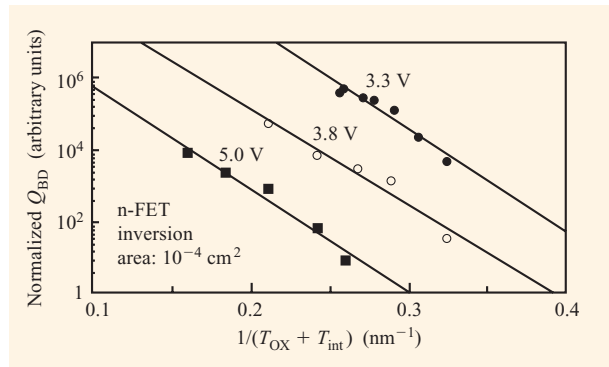


Figure 3

Charge-to-breakdown, Q_{BD} , vs. thickness at different voltages. To allow such a wide range of T_{OX} at a fixed voltage, we have used either long-term stress or smaller-area structures, then scaled to a reference area of 10^{-4} cm^2 using the Weibull slopes of Figure 2. Reprinted with permission from [23].

Recently it has been suggested that N_{BD} is dependent on voltage and temperature [17, 18]. We have conducted direct T_{BD} (Q_{BD}) measurements and found that Weibull slopes are independent of stress voltage, temperature, and polarity over a wide range of these variables [19, 20]. These results show that the simple geometric picture proposed in the percolation [12, 13] and analytical models [14] is consistent with the experimental statistics, using Q_{BD} (T_{BD}) as the breakdown variable. Moreover, it was recently shown that SILC-based measurements cannot be reliably used as a measure of the critical defect density, N_{BD} , because they fail to capture the correct thickness dependence of Q_{BD} Weibull slopes [20]. In addition, there is currently no detailed understanding of the microscopic mechanisms that cause breakdown defects. Thus, introduction of voltage or temperature dependencies into the geometrical picture of breakdown statistics must be made on an *ad hoc* basis.

4. Voltage dependence of oxide breakdown

Another key to accurate assessment of oxide reliability is the voltage-acceleration factor. Several important studies have been carried out [4, 5, 7–30] on this subject. Advances in the anode hole-injection model, the hydrogen model, and the thermochemical model have been made [7–10, 21–24]. Very large voltage-acceleration factors have been reported [17, 26–28] for ultrathin oxides, and voltage dependence of the voltage-acceleration factor has recently been demonstrated [17, 26]. Historically, for thick oxides (>6 nm) stressed at high voltages in the nonballistic Fowler–Nordheim (FN) regime, oxide breakdown has been considered to be field-driven, because the electric field can affect the energy of electrons exiting from oxides [7]. For

ultrathin oxides stressed or operated in ballistic FN-tunneling and direct-tunneling regimes, electron energy at the anode can directly affect the release of (secondary) species at the anode [7, 25]. Thus, electron energy, or gate voltage, was first proposed to be the controlling variable in the oxide-breakdown process [25]. However, more detailed work is required to further illuminate the role of carrier energy and oxide field in defect-generation rate and oxide breakdown. In the following discussion, gate voltage is considered to be an independent variable.

Figure 4(a) shows T_{BD} vs. gate voltage for several values of T_{OX} . Defining the (local) voltage-acceleration factor as $\gamma(V) = -\partial \ln T_{BD} / \partial V$, to be consistent with the widely used exponential law, $T_{BD} \sim \exp(-\gamma V)$, the data in Figure 4(a) reveal that the voltage acceleration appears to increase for thinner oxides; however, a complication is presented by the fact that thinner oxides are stressed at lower voltages for practical reasons. In Figure 4(b), experimentally obtained T_{BD} values for 1.7-, 2.3-, and

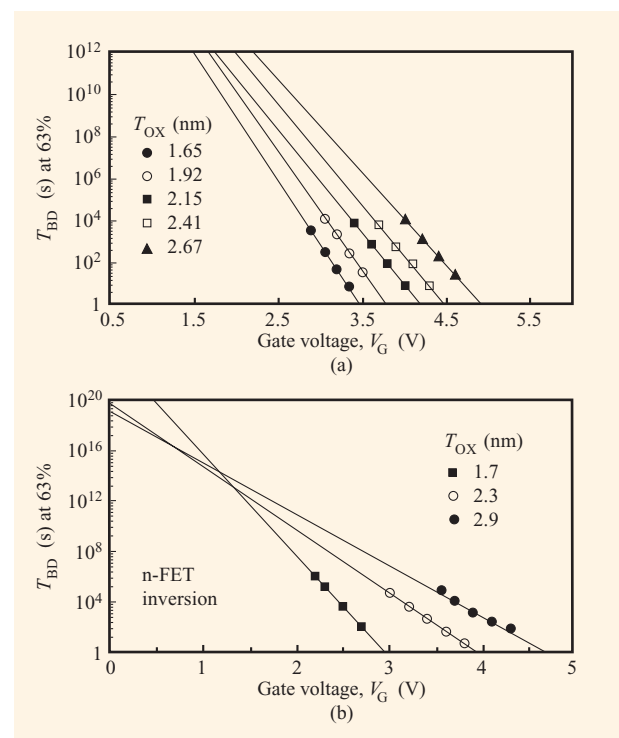


Figure 4

(a) T_{63} vs. V_G for thinner oxides with p^+ poly/SiO₂/Si(n) capacitors under substrate injection. Reprinted with permission from [26]; ©2000 IEEE. (b) Comparison of projected T_{BD} from stress data using voltage dependence of an exponential law for 1.7-, 2.3-, and 2.9-nm oxides at the same area. The T_{BD} data for 1.7-nm oxides were translated from those for smaller areas using the Weibull slopes of Figure 2. Reprinted with permission from [10]; ©2001 IEEE.

2.9-nm oxides stressed at low voltages are compared to predictions of T_{BD} based on the exponential law, with each oxide thickness independently fitted with a constant acceleration factor. This comparison suggests that if the exponential law of voltage dependence were to remain true over the entire voltage range, including voltages as low as use conditions, the lifetime of thinner oxides would ultimately exceed that of thick oxides, as depicted in Figure 4(b). This is incompatible with the concept of the critical defect density decreasing with reduced T_{OX} and suggests that the apparent thickness dependence of voltage acceleration may be an artifact resulting from the necessity of collecting T_{BD} at lower voltages for thinner oxides.

To demonstrate the voltage dependence of the voltage acceleration factor, two independent experimental methods were employed: area scaling and long-term stress. **Figure 5(a)** shows T_{BD} data measured on four oxide areas for 2.67-nm oxide, on a semi-log plot. It is clear that a simple exponential fit to the data in Figure 5(a) gives nonparallel slopes (i.e., voltage-acceleration factors are different) when a wide range of voltages are used. Similarly, if the exponential law for each fit were to remain true over this range of voltage, the extrapolated lines would eventually cross over at low voltages. This in turn suggests that the lifetime of the smaller-area structures would be shorter than that of the larger-area structures, and provides yet another unphysical result, contradicting the weakest-link nature of oxide breakdown. On the basis of these results, it is readily concluded that the exponential law for T_{BD} (Q_{BD}) voltage dependence cannot hold over a wide range of voltage. In particular, one should expect that the voltage-acceleration factor must increase with decreasing voltage.

As an independent test of this T_{BD} voltage dependence, T_{BD} measurements at lower voltages were conducted for much longer stress times, over one-half year for 1.7-nm oxides, as shown in **Figure 5(b)**, and down to stress voltages of ~ 2 V. The T_{BD} data (filled circles) for large-area capacitors in Figure 5(b) were normalized using Equation (2) with the β value taken from Figure 2. A fit was made to only the higher-voltage data, typically acquired from wafer-level stress, and then the fitted curve was extrapolated to low voltages. It is clear that the low-voltage T_{BD} data deviate from the extrapolation made from the high-voltage data. Moreover, in Figure 5(b), a high-voltage-only fit gives a voltage acceleration, $\gamma_H = 15.2$, while the low-voltage-only fit yields $\gamma_L = 18.6$. The error bars for voltage-acceleration factors are ± 0.6 . The γ values measured at high and low voltages are statistically different. These results provide direct experimental evidence of a change of voltage-acceleration factors with voltage. Not only do the experimental results of these two independent methods support the claim that voltage-acceleration factors are voltage-dependent, but such a

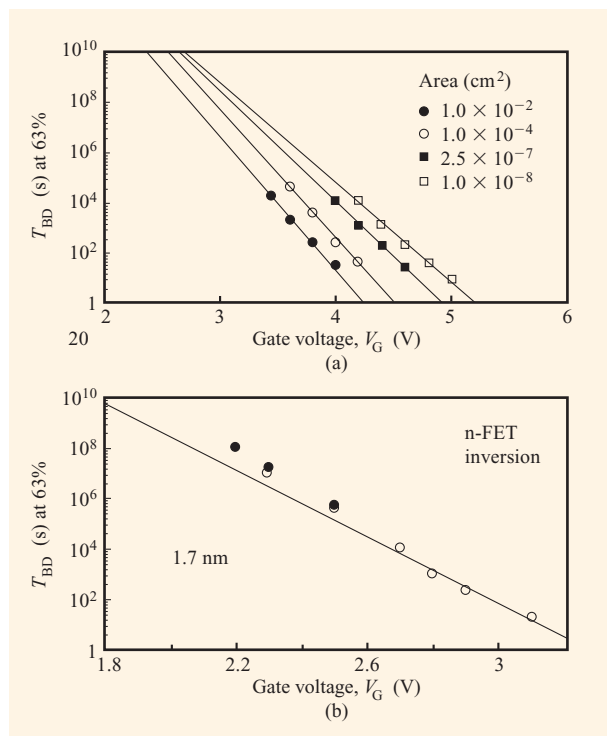


Figure 5

(a) T_{BD} vs. V_G for four different areas with 2.67-nm oxide using p^+ poly/Si(n) structures. (b) Long-term module stress results on n-FETs having two different areas. Parts (a) and (b) reprinted with permission from [26]; ©2000 IEEE.

dependence is also consistent with the concept of electron energy or gate voltage [25] as the primary breakdown variable, as was previously demonstrated using independent (polysilicon) gate-electrode-depletion experiments [29, 30]. **Figure 6** displays the voltage-acceleration factor as a function of voltage across several CMOS technologies for a wide range of oxide thickness from 1.2 nm to 5.0 nm. These voltage-acceleration factors are derived from semi-log plots such as those of Figure 5(a). Because the corresponding voltages along the horizontal axis are the average values of the voltages, these voltage-acceleration factors are only approximations of the true differential voltage-acceleration factors. The extensive volume of experimental data shown in Figure 6 demonstrates the universality and predictive capability of the power-law relation for voltage acceleration [26].

5. Defect-generation rate

It is widely accepted that the oxide-breakdown process related to the generation of defects as a result of stress proceeds in two steps:

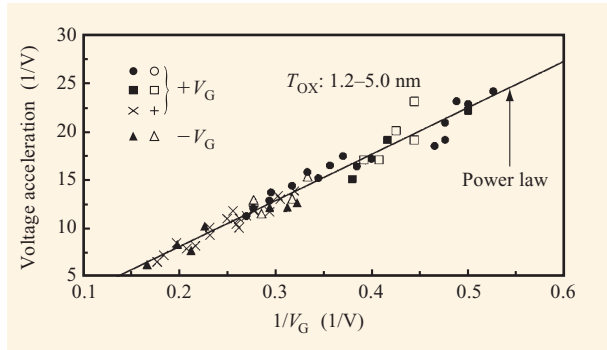


Figure 6

Comparison of the measured T_{BD} voltage-acceleration γ values with the predictions of T_{BD} power law. The experimental data represent three sets of n-FET devices from 1.2 nm to 5.0 nm and the p⁺ poly/SiO₂/Si(n) capacitors from 1.6 nm to 5.0 nm. Reprinted with permission from [26]; ©2000 IEEE.

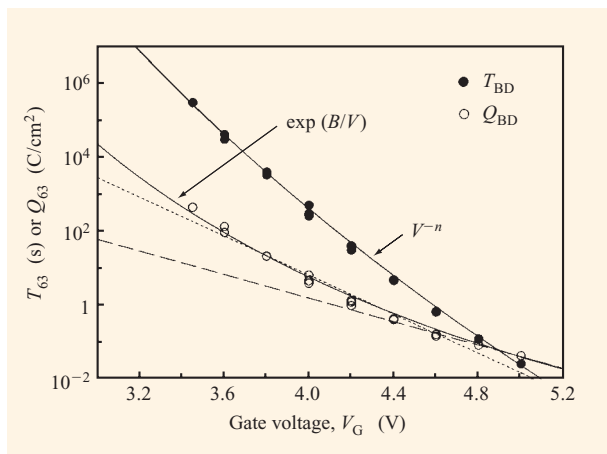


Figure 7

Normalized Q_{BD} and T_{BD} vs. V_G for 2.67-nm oxides at 140°C. The corresponding T_{BD} is also included. The data were normalized using β values in Figure 1. The solid curves represent the fitted curves using $\exp(B/V)$ and V^{-n} for Q_{BD} and T_{BD} , respectively, with $B = 100$ V and $n = 44$. The dotted curve is for a least-squares fit of $\ln Q_{BD}$ vs. V , whereas the dashed curve represents a least-squares fit of only high-voltage data points. Reprinted with permission from [10]; ©2001 IEEE.

1. A statistical breakdown process associated with the critical defect density, i.e., the triggering of breakdown by the density of generated defects at breakdown.
2. A deterministic breakdown process associated with a defect-generation rate, i.e., the evolution of the average density of defects with time or injected charge as a function of stress conditions such as voltage and temperature [12–14].

Assuming that there is a linear dependence of defect generation on electron fluence [4], a general framework for describing these two processes [4, 23] is as follows:

$$Q_{BD}(T_{OX}, A_{OX}, V, T) = \frac{N_{BD}(T_{OX}, A_{OX})}{\xi(V, T)}, \quad (3)$$

where N_{BD} is the critical defect density at breakdown, reflecting the statistical nature of oxide breakdown, as discussed above. This first process appears to be well understood, based on the success of percolation and analytical models in explaining the thickness and area dependences of breakdown statistics such as Weibull slopes, as discussed in the previous section. The term $\xi(V, T)$ describes the defect-generation rate at a given stress condition. Unfortunately, the physical nature of the second process associated with $\xi(V, T)$ is still under investigation and is being debated.

Considering the preceding experimental evidence of voltage-independent and temperature-independent Weibull slopes [19, 20], and in the absence of microscopic information on possible voltage and temperature dependence of N_{BD} , the defect-generation rate is presumed to be dependent exclusively on voltage and temperature. With this approach, and using Equation (1), the Q_{BD} voltage dependence can be used to measure the defect-generation rate, $\xi(V)$. **Figure 7** shows the measured Q_{BD} and T_{BD} data vs. gate voltage using different-area capacitors. The defect-generation rate, as measured by $Q_{BD}(V)$, follows an $\exp(B/V)$ dependence, indicating a much-reduced defect-generation rate with reduced voltage. It can be shown analytically, by considering $Q_{BD} \approx T_{BD} J_0$, where J_0 is the tunneling current, that the power-law T_{BD} voltage dependence is self-consistent with the $\exp(B/V)$ voltage dependence for Q_{BD} . This connection gives rise to the large value of the exponent (~ 40) found in the power-law relation [26]. Note that the parameter B can be voltage-dependent [10, 26].

6. Temperature dependence of oxide breakdown

Several experimental observations regarding both temperature and voltage dependence of ultrathin-oxide breakdown have been made. First, a very strong T_{BD}/Q_{BD} temperature dependence has been found for thin oxides as compared to thick oxides [18, 31–34]. This stronger temperature dependence for thinner oxides was first interpreted as a thickness effect in [18] and supported by other reports [31–34]. Second, a non-Arrhenius temperature dependence of oxide breakdown has also been reported for ultrathin oxides [18, 31–34]. Third, it has also been reported that the voltage-acceleration factor

remains unchanged over a range of temperatures (30°C to 200°C) but within a fixed T_{BD} or Q_{BD} window [33–35]. These results may seem confusing or even contradictory. Recently, these observations were experimentally resolved in the context of voltage-dependent voltage acceleration together with a unified, coherent global picture which is discussed below [19].

Normalized T_{BD} is plotted as a function of voltage for different temperatures in **Figure 8**. T_{BD} data from different-area capacitors were normalized to equivalent areas using the Weibull slopes from Figure 2. The data from different oxide thicknesses were then further normalized to isolate the uniquely voltage-dependent behavior, using a thickness-scaling law [14]. This analysis reveals some very important findings:

1. At a fixed temperature, the voltage-acceleration factor increases with decreasing voltage, as discussed in the previous section.
2. At lower temperatures, T_{BD} increases at a faster rate with decreasing voltage than it does at higher temperatures.
3. About a fixed voltage, the voltage-acceleration factor is larger at lower temperature than at higher temperature, while about a fixed T_{BD} , the voltage-acceleration factor remains unchanged with temperature.
4. At lower voltages, a more pronounced temperature dependence is observed than at higher voltages.

The results in Figure 8 also suggest that the strong Q_{BD} (T_{BD}) temperature dependence found in ultrathin oxides is not due to the small value of T_{OX} , as previously hypothesized [18, 31–34], but rather due to the low stress voltage. This has recently been verified by performing stresses at different voltages with T_{OX} fixed, using capacitors of various areas to provide the needed range of data [19]. These results demonstrate that additional reliability margin can be realized by accurately taking into account the effects of temperature when making reliability projections. In particular, the increase in T_{BD} and in the voltage-acceleration factor at low voltage provides significant relief in reliability projections, enabling the use of thinner oxides than conventionally projected.

7. Oxide breakdown vs. device circuit failure

Oxide failure is marked by a sudden jump in leakage current across the oxide film as a result of stress. For ultrathin oxides (<6.0 nm), the post-breakdown current can vary by many orders of magnitude depending on stress conditions, oxide thickness, and the external circuit environment. Breakdown is typically classified as being either “hard” or “soft” depending on whether the post-breakdown conduction is respectively at the high end or the low end of the observed distribution. Currently, there is

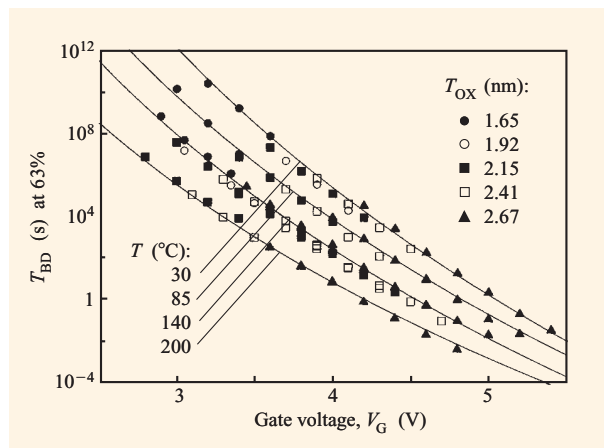


Figure 8

Normalized T_{BD} vs. V_G at different temperatures: 30°C, 85°C, 140°C, and 200°C. Area scaling and thickness scaling have been applied and are discussed in the text. Reprinted with permission from [10]; ©2001 IEEE.

considerable activity and debate concerning the proper interpretation of so-called soft- and hard-breakdown phenomena observed in ultrathin oxides.

Substantial progress has been made toward understanding of soft- and hard-breakdown modes and their implications for device operation and circuit function [36–49], although a universally accepted methodology has not yet emerged. Naturally, several researchers have raised the question of exactly what the best criteria are for categorizing a change in oxide electrical conduction as a breakdown event [42–49]. Several detailed studies of the effects of oxide breakdown on device failure show that oxide breakdown in the drain or source region is more detrimental to device operation than oxide breakdown in the channel region [43, 44]. On the other hand, it is found that the severity of oxide breakdown, defined as the ratio of the prevalence of soft breakdown to hard breakdown, is voltage-dependent, with soft-breakdown events becoming relatively more probable as the voltage is reduced [42, 45–49].

The voltage dependence of the fail mode (e.g., soft vs. hard) is now relatively well understood [39, 42, 46–48]. When the power dissipation through the percolation path exceeds a threshold, a hard-breakdown event is reached [45–48]. Subsequent study reveals that hard breakdown can occur only when the total amount of energy exceeds some energy threshold [47]. The experimental data on soft and hard breakdown compare satisfactorily with predictions based on such an energy model [46, 47], and a new methodology which takes this effect into account has been proposed [47]. In a circuit environment, transistors

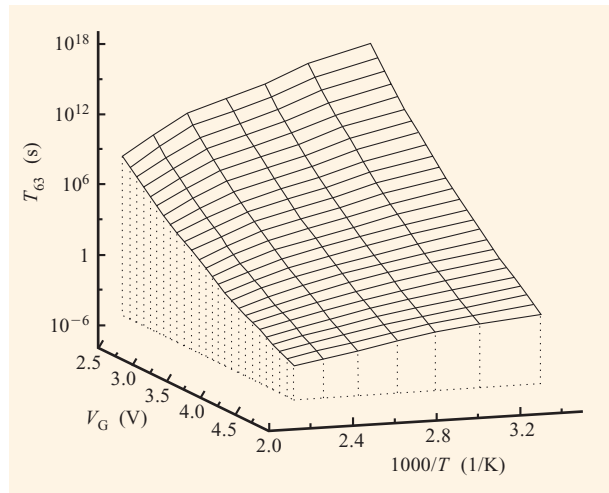


Figure 9

Three-dimensional plot for time-to-breakdown vs. voltage and temperature constructed from the simple empirical principles of Equations (4) through (6). Reprinted with permission from [10]; ©2001 IEEE.

typically operate in a current-limited mode [48], that is, there are electrical elements in series with a gate-discharge path which will limit the maximum current in the event of a gate-oxide failure. The value of the current to which a breakdown is limited (by the circuit) significantly affects the post-breakdown leakage currents [48], and this in turn leads to a reduction of the oxide-failure hazard. This was experimentally demonstrated in the case of the ring oscillator circuit [49], where functionality was maintained even after a hard-breakdown event. Thus, these studies suggest that in a realistic circuit environment, the circuit functionality may be much more resilient with respect to oxide breakdown than was previously assumed in assessing the oxide reliability of integrated circuits.

Even if the occurrence of oxide breakdown does not destroy device operation, however, it is premature to claim that devices will remain completely functional. Erratic behavior of devices and a progressive degradation after (initial) oxide breakdown have been reported [11]. Significantly more research effort is required in order to understand the voltage and temperature dependence of the post-breakdown events, the eventual device breakdown, and the statistical distribution of device breakdown events. A complete methodology is needed to provide a predictive model for the ultimate impact on circuit operation, including the appropriate criteria for circuit failure as a result of oxide breakdown. Once a complete methodology is developed, it is anticipated that

additional reliability margin will be realized for ultrathin oxides.

8. Reliability outlook

On the basis of our experimental investigation, a global view of time-to-breakdown for ultrathin oxides is now detailed. While the fundamental physical mechanisms behind the observed reduction in defect generation with decreasing voltages remain unknown, all of the experimental data presented in this paper are accurately captured by two simple empirical principles:

$$\frac{V}{T_{BD}} \frac{\partial T_{BD}}{\partial V} = \text{const} = n(T) \quad (4)$$

and

$$\frac{d}{dT} \left(\frac{1}{T_{BD}} \frac{\partial T_{BD}}{\partial V} \right)_{T_{BD}} = 0, \quad (5)$$

where V and T respectively represent the stress voltage and temperature. Equation (4) describes a power-law dependence of T_{BD} on voltage and accurately captures the entirety of the experimental data presented in this work, without qualification. Equation (5) describes the (local) voltage acceleration at a fixed value of T_{BD} , irrespective of temperature or T_{OX} ; i.e., regardless of the combination of the variables used to arrive at a given T_{BD} , $\gamma(V)$ will have the same value! A complete picture for time-to-breakdown, in both the voltage and temperature domains, can be developed on the basis of these two simple principles, with the addition of a third relationship that describes T_{BD} temperature dependence for a given voltage:

$$T_{BD} = T_{BD0}(V) \exp\left(\frac{a(V)}{T} + \frac{b(V)}{T^2}\right), \quad (6)$$

where the pre-factor, T_{BD0} , and the coefficients, a and b , are voltage-dependent. The second-order term, b/T^2 , empirically inserts non-Arrhenius temperature effects. These empirical principles can now be readily used to make reliability projections for ultrathin gate oxides.

Figure 9 shows a three-dimensional plot for T_{BD} vs. voltage and temperature constructed from these principles, using the experimental data presented in this work to determine the required constants.

Figure 10 displays the maximum allowable supply (gate) voltage as a function of oxide thickness. This reliability projection is based on a reliability requirement of 100-ppm failure rate and 0.1 cm² of gate area [4]. The dashed curves for sub-1.0-nm oxides are extended from the projected maximum voltages for oxide thickness ≥ 1.0 nm assuming the same thickness scaling law and the same T_{BD} voltage dependence as discussed below. It is evident from **Figure 10** that gate-oxide reliability will not inhibit realization of the ITRS roadmap. Thus, the tunneling

current, as a circuit and product design issue only, remains the limiting factor for scaling gate-oxide thickness until an alternative gate dielectric becomes available.

To assess the limitations of our reliability projections, in **Figure 11** we plot normalized lifetime as a function of voltage for experimental results down to 1.8 V as well as reliability projections from ~ 2 V to ~ 1 V. Above ~ 2 V, the universal curve spanning ~ 20 orders of magnitude in T_{BD} is obtained by applying the area-scaling and thickness-scaling relationships [14] to the experimental data, as discussed in Section 3. Below ~ 2 V, several possible scenarios may occur, as indicated by the dashed curve and the shaded area. The dashed curve illustrates the consequences of the (*ad hoc*) assumption that the voltage-acceleration factor remains constant below 2 V. The shaded area represents many possible reliability projections for new (currently unknown) mechanisms. On the basis of SILC measurements, an increase in defect-generation rate described by a sigmoidal behavior has been reported [17]. However, the SILC-based measurements appear to deviate from experimental T_{BD} (Q_{BD}) when a quantitative comparison is made [4, 17]. Therefore, this possibility appears to be at odds with experimental observations. Arrows are used to indicate the potential for improvement in reliability projections should a complete methodology for soft breakdown of devices and for circuit failure criteria be realized.

9. Conclusions

We have conducted a systematic and extensive investigation of ultrathin-oxide breakdown based on a highly accurate experimental methodology. The thickness dependence of Q_{BD}/T_{BD} and Weibull slopes has been studied in the context of a decrease in the critical defect density with decreasing oxide thickness. Weibull slopes are found to be independent of voltage and temperature, supporting the geometrical picture proposed by percolation- and cell-based models. Direct evidence of voltage-dependent voltage acceleration has been established using two experimental, independent methods. This voltage-dependent voltage acceleration, ascribed to a power-law behavior of T_{BD} with voltage, was found to be universal across a wide range of oxide thickness, temperature, and voltage. Q_{BD}/T_{BD} temperature dependence and constant Weibull slope are found to be consistent with voltage-dependent voltage acceleration; thus, a global and coherent picture emerges. This large-volume experimental database, together with an accurate experimental methodology, enables timely reliability projections for advanced CMOS technologies. Provided that no new mechanisms arise for oxides thinner than 1.2 nm, it is concluded that silicon-dioxide-based gate dielectric can be reliably extended beyond the 100-nm node for CMOS technologies.

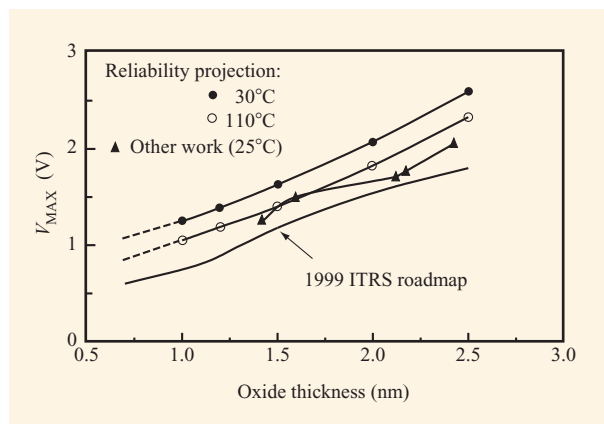


Figure 10

Projected supply voltage as a function of oxide thickness at two temperatures for 100 ppm failure rate and 0.1 cm^2 gate area. The dashed curves are extended from the projected maximum voltages for oxide thickness ≥ 1 nm. Projections from other work are taken from [27].

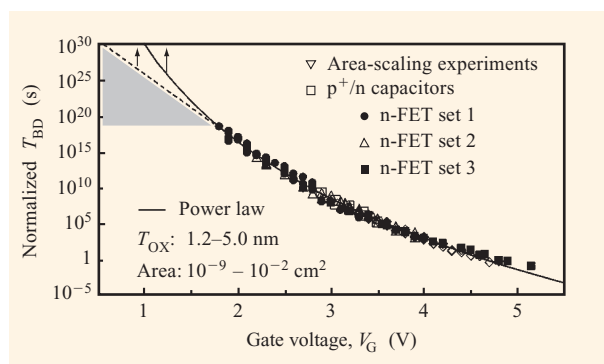


Figure 11

Normalized lifetime as a function of voltage for experimental data and reliability projection below 2 V with several different scenarios discussed in the text. The dashed curve is the reliability projection assuming a constant voltage-acceleration factor below 2 V. The shaded area represents possible reliability projections due to unknown new mechanisms, as discussed in the text. The arrows indicate the improvement in the future reliability projection if a complete methodology for soft breakdown of devices and circuits is realized.

Acknowledgments

We are grateful to J. Aitken and our management for their constant support throughout this work over many years. It is our great pleasure and honor to acknowledge the stimulating and insightful discussion with Prof. J. Suñé on many aspects of oxide-breakdown phenomena. We are

honored to acknowledge the technical advice and stimulating discussions of R. Vollertsen over many years of working in this field. We sincerely thank J. McKenna, R. Dufresne, C. Montrose, D. Brochu, and B. Morse for their constant technical support during this work.

References

1. Y. Taur and E. Nowak, "CMOS Devices Below 0.1 μm ; How High Will Performance Go?," *IEDM Tech. Digest*, pp. 215–218 (1997).
2. D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, "The Electronic Structure at the Atomic Scale of Ultrathin Gate Oxides," *Nature* **399**, 758–761 (1999).
3. C. Hu, "Gate Oxide Scaling Limits and Projections," *IEDM Tech. Digest*, pp. 96–99 (1996).
4. J. H. Stathis and D. J. DiMaria, "Reliability Projection for Ultra-Thin Oxides at Low Voltage," *IEDM Tech. Digest*, pp. 167–170 (1998).
5. R. Degraeve, B. Kaczer, and G. Groeseneken, "Reliability: A Possible Showstopper for Oxide Thickness Scaling?," *Semicond. Sci. Technol.* **15**, 436–444 (2000).
6. *International Technology Roadmap for Semiconductors*, 1999 Edition, Semiconductor Industry Association, 4300 Stevens Suite Boulevard, Suite 271, San Jose, CA 95129.
7. D. J. DiMaria, E. Cartier, and D. Arnold, "Impact Ionization, Trap Creation, Degradation, and Breakdown in Silicon Dioxide Films on Silicon," *J. Appl. Phys.* **73**, 3367–3384 (1993).
8. I. C. Chen, S. E. Holland, K. K. Young, C. Chang, and C. Hu, "Substrate Hole Current and Oxide Breakdown," *Appl. Phys. Lett.* **74**, 669–671 (1986).
9. J. W. McPherson and H. C. Mogul, "Underlying Physics of the Thermo-Chemical E Model in Describing Low-Field Time-Dependent Breakdown in SiO_2 Thin Films," *J. Appl. Phys.* **84**, 1513–1523 (1998).
10. E. Wu, E. Nowak, A. Vayshenker, J. McKenna, D. Harmon, and R. Vollertsen, "New Global Insight in Ultra-Thin Oxide Reliability Using Accurate Experimental Methodology and Comprehensive Database," *IEEE Trans. Device & Mater. Reliabil.* **1**, 69–80 (2001).
11. E. Y. Wu, J. H. Stathis, and L.-K. Han, "Ultra-Thin Oxide Reliability for ULSI Applications," *Semicond. Sci. Technol.* **15**, 425–435 (2000).
12. R. Degraeve, G. Groeseneken, R. Bellens, J. Ogier, M. Depas, P. J. Roussel, and H. E. Maes, "A Consistent Model for the Thickness Dependence of Intrinsic Breakdown in Ultra-Thin Oxides," *IEDM Tech. Digest*, pp. 866–869 (1995).
13. J. H. Stathis, "Percolation Models for Gate Oxide Breakdown," *J. Appl. Phys.* **86**, 904–911 (1998).
14. J. Suñé, "New Physics-Based Analytic Approach to the Thin Oxide Breakdown Statistics," *IEEE Electron Device Lett.* **22**, 296–298 (2001).
15. 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* **185**, 347–362 (1990).
16. D. J. DiMaria and J. H. Stathis, "Explanation of Oxide Thickness Dependence of Breakdown Characteristics of Metal-Oxide-Semiconductor Structures," *Appl. Phys. Lett.* **70**, 2708–2710 (1997).
17. J. H. Stathis, A. Vayshenker, P. Varekamp, E. Y. Wu, C. Montrose, J. McKenna, D. J. DiMaria, L.-K. Han, E. Cartier, R. Wachnik, and B. P. Linder, "Breakdown Measurements of Ultra-Thin SiO_2 at Low Voltage," *Symposium on VLSI Technology, Digest of Technical Papers*, 2000, pp. 94–95.
18. D. J. DiMaria and J. H. Stathis, "Non-Arrhenius Temperature Dependence of Reliability in Ultrathin Silicon Dioxide Films," *Appl. Phys. Lett.* **74**, 1752–1754 (1999).
19. E. Wu, J. Suñé, W. Lai, E. Nowak, J. McKenna, A. Vayshenker, and D. Harmon, "Interplay of Voltage and Temperature Acceleration of Oxide Breakdown for Ultra-Thin Oxides," *Proceedings of the Conference on Insulating Films on Semiconductors (INFOS)*, Udine, Italy, 2001, pp. 13–14.
20. E. Wu, J. Suñé, E. Nowak, W. Lai, and J. McKenna, "Weibull Slopes, Critical Defect Density, and the Validity of SILC Measurements," *IEDM Tech. Digest*, pp. 125–128 (2001).
21. M. Alam, J. Budd, and A. Ghetti, "Field Acceleration for Oxide Breakdown—Can an Accurate Anode Hole Injection Model Resolve the E vs. $1/E$ Controversy?," *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 21–26.
22. J. Suñé and E. Wu, "A New Quantitative Hydrogen-Based Model for Ultra-Thin Oxide Breakdown," *Symposium on VLSI Technology, Digest of Technical Papers*, 2001.
23. J. Suñé and E. Wu, "Modeling Breakdown and Breakdown Statistics of Ultra-Thin SiO_2 Gate Oxides," *Microelectron. Eng.* **59**, 149–153 (2001).
24. J. W. McPherson and R. B. Khamankar, "Molecular Model for Intrinsic Time-Dependent Dielectric Breakdown in SiO_2 Dielectrics and Reliability Implications for Hyper-Thin Gate Oxide," *Semicond. Sci. Technol.* **15**, 462–470 (2000).
25. D. J. DiMaria, "Dependence on Gate Work Function of Oxide Charging, Defect Generation, and Hole Currents in Metal-Oxide-Semiconductor Structures," *J. Appl. Phys.* **81**, 3220–3226 (1997).
26. E. Wu, J. Aitken, E. Nowak, A. Vayshenker, P. Varekamp, G. Hueckel, J. McKenna, D. Harmon, L.-K. Han, C. Montrose, R. Dufresne, and R.-P. Vollertsen, "Voltage-Dependent Voltage Acceleration of Oxide Breakdown for Ultra-Thin Oxides," *IEDM Tech. Digest*, pp. 54–57 (2000).
27. B. Weir, M. Alam, J. D. Bude, P. J. Silverman, A. Ghetti, F. Baumann, P. Diodato, D. Monroe, T. Sorsch, G. L. Timp, Y. Ma, M. Brown, A. Hamad, D. Hwang, and P. Mason, "Gate Oxide Reliability Projection to the Sub-2 nm Regime," *Semicond. Sci. Technol.* **15**, 455–461 (2000).
28. M. Takayanagi, S. Takagi, and Y. Toyoshima, "Experimental Study of Gate Voltage Scaling for TDDB Under Direct Tunneling Regime," *Proceedings of the International Reliability Physics Symposium*, 2001, pp. 380–385.
29. P. Nicollian, W. R. Hunter, and J. C. Hu, "Experimental Evidence for Voltage Driven Breakdown Models in Ultra-Thin Gate Oxides," *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 7–15.
30. J. McKenna, E. Y. Wu, and S.-H. Lo, "Tunneling Current Characteristics and Oxide Breakdown in p+poly Poly Gate PFET Capacitors," *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 16–20.
31. B. Kaczer, R. Degraeve, N. Pangon, and G. Groeseneken, "The Influence of Elevated Temperature on Degradation and Lifetime Predication of Thin Silicon-Dioxide Films," *IEEE Trans. Electron Devices* **47**, 1514–1521 (2000).
32. R. Degraeve, G. Groeseneken, B. Kaczer, and N. Pangon, "Temperature Acceleration of Oxide Breakdown and Its Impact on Ultra-Thin Gate Oxide Reliability," *Symposium on VLSI Technology, Digest of Technical Papers*, 1999, pp. 59–60.
33. E. Wu, D. Harmon, and L.-K. Han, "Interrelationship of Voltage and Temperature Dependence of Oxide Breakdown for Ultrathin Oxides," *IEEE Electron Device Lett.* **21**, 362–364 (2000).
34. J. Suehle, E. Vogel, B. Wang, and J. Bernstein, "Temperature Dependence of Soft Breakdown and

Wear-Out in Sub-3nm SiO₂ Films,” *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 33–39.

Received October 15, 2001; accepted for publication January 18, 2002

35. T. Pompl, H. Wurzer, M. Kerber, and I. Eisele, “Investigation of Ultra-Thin Gate Oxide Reliability Behavior by Separate Characterization of Soft Breakdown and Hard Breakdown,” *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 40–47.
36. K. Okada, S. Kawasaki, and Y. Hirofujii, “New Experimental Findings on Stress Induced Leakage Current of Ultra-Thin Oxide Silicon Oxides,” *Extended Abstracts of the 1994 International Conference on Solid State Devices and Materials*, 1994, pp. 565–567.
37. S.-H. Lee, B.-J. Cho, J.-C. Kim, and S.-H. Choi, “Quasi-Breakdown of Ultra-Thin Gate Oxide Under High Field Stress,” *IEDM Tech. Digest*, pp. 605–607 (1994).
38. M. Depas, T. Nigam, and M. M. Heyns, “Soft Breakdown of Ultra-Thin Gate Oxide Layers,” *IEEE Trans. Electron Devices* **43**, 2499–2504 (1996).
39. S. Lombardo, F. Crupi, A. La Magna, C. Spinella, A. Terrasi, A. La Mantia, and B. Neri, “Electrical and Thermal Transient During Dielectric Breakdown of Thin Oxides in Metal-SiO₂ Capacitors,” *J. Appl. Phys.* **84**, 472–479 (1998).
40. J. Suñé and E. Miranda, “Post Soft Breakdown Conduction in SiO₂ Gate Oxides,” *IEDM Tech. Digest*, pp. 533–536 (2000).
41. H. Satake and A. Toriumi, “SiO₂ Dielectric Breakdown Mechanism Studied by the Post Breakdown Resistance Statistics,” *IEEE Trans. Electron Devices* **47**, 741–745 (2000).
42. B. E. Wier, P. J. Silverman, D. Monroe, K. S. Krisch, M. A. Alam, G. B. Alers, T. Sorsch, G. L. Timp, F. Baumann, C. T. Liu, Y. Ma, and D. Huang, “Ultra-Thin Gate Dielectrics: They Break Down, But Do They Fail?,” *IEDM Tech. Digest*, pp. 73–76 (1997).
43. E. Wu, E. Nowak, J. Aitken, W. Abadeer, L.-K. Han, and S.-H. Lo, “Structure Dependence of Dielectric Breakdown in Ultra-Thin Gate Oxides, and Its Relationship to Soft Breakdown Modes and Device Failure,” *IEDM Tech. Digest*, pp. 187–190 (1998).
44. R. Degraeve, B. Kaczer, A. Keersgieter, and G. Groeseneken, “Relation Between Breakdown Mode and Breakdown Location in Short Channel NMOSFETs and Its Impact on Reliability Specifications,” *Proceedings of the International Reliability Physics Symposium*, 2001, pp. 360–366.
45. S. Bruyere, E. Vincenet, and G. Ghibaudo, “Quasi-Breakdown in Ultra-Thin SiO₂ Films: Occurrence Characterization and Reliability Assessment Methodology,” *Proceedings of the International Reliability Physics Symposium*, 2000, pp. 48–54.
46. M. A. Alam, B. Weir, P. Silverman, Y. Ma, and D. Hwang, “The Statistical Distribution of Percolation Resistance as a Probe into the Mechanics of Ultra-Thin Oxides Breakdown,” *IEDM Tech. Digest*, pp. 529–532 (1999).
47. J. Suñé, E. Y. Wu, D. Jiménez, R. P. Wollertsen, and E. Miranda, “Understanding Soft and Hard Breakdown Statistics, Prevalence Ratios and Energy Dissipation During Breakdown Runaway,” *IEDM Tech. Digest*, pp. 117–120 (2001).
48. B. Linder, J. H. Stathis, R. A. Wachnik, W. E. Wu, S. A. Cohen, A. Ray, and A. Vayshenker, “Gate Oxide Breakdown Under Current Limited Constant Voltage Stress,” *Symposium on VLSI Technology, Digest of Technical Papers*, 2000, pp. 214–215.
49. B. Kaczer, R. Degraeve, G. Groeseneken, M. Rasras, S. Kubicek, E. Vandamme, and G. Badenes, “Impact on MOSFET Oxide Breakdown on Digital Circuit Operation and Reliability,” *IEDM Tech. Digest*, pp. 553–556 (2000).

Ernest Y. Wu *IBM Microelectronics Division, 1000 River Road, Essex Junction, Vermont 05452 (eywu@us.ibm.com).* Dr. Wu is a Senior Engineer in the Technology Reliability Department of the IBM Microelectronics Division in Essex Junction, Vermont. He received a B.S. degree in engineering physics and M.S. and Ph.D. degrees in physics from the University of Kansas in 1984, 1986, and 1989, respectively. Dr. Wu joined the IBM Storage Products Division in 1989. In 1994, he transferred to the IBM Microelectronics Division in Essex Junction, Vermont. He is responsible for technology qualification of ultrathin gate oxides. His research interests include reliability of ultrathin oxides, device reliability, carrier transport, and magnetics. Dr. Wu has served on the Device Dielectric Committee as a co-chair for the 2000 International Reliability Physics Symposium (IRPS). He was a member of the CMOS and Interconnect Reliability Committee of the International Electron Devices Meeting (IEDM) for 1999 and 2000. Dr. Wu has authored or co-authored more than 40 technical journal and conference papers.

Edward J. Nowak *IBM Microelectronics Division, 1000 River Road, Essex Junction, Vermont 05452 (ejnowak@us.ibm.com).* Dr. Nowak is a Senior Technical Staff Member in the CMOS Process Development Group. He received a B.S. degree in physics from M.I.T. in 1973, and M.S. and Ph.D. degrees in physics from the University of Maryland in 1974 and 1978, respectively. Following postdoctoral research at New York University, he joined IBM in Essex Junction in 1981 to work on memory device and process development. In 1985 he turned to high-speed CMOS logic devices spanning 1- μm to 50-nm device designs. Dr. Nowak's current research interests are in double-gate CMOS and low-power CMOS. He is an author or coauthor of more than 50 U.S. patents and more than 40 technical papers and a book in the areas of VLSI technology and design. Dr. Nowak is a member of the Institute of Electrical and Electronics Engineers and the American Physical Society.

Alex Vayshenker *IBM Microelectronics Division, 1580 Route 52, Hopewell Junction, New York 12533 (vaysha@us.ibm.com).* Mr. Vayshenker received a B.S. degree in electro-optical engineering from Moscow University in 1978. Since joining IBM in 1984, he has held a variety of engineering positions in semiconductor development. He is currently an Advisory Engineer, working on reliability of ultrathin dielectrics. In 2000, Mr. Vayshenker received an IBM Outstanding Technical Achievement Award for his work on SiGe HBT technology. He has coauthored several papers on gate dielectric reliability and SiGe technology.

Wing L. Lai *IBM Microelectronics Division, 1000 River Road, Essex Junction, Vermont 05452 (winglai@us.ibm.com).* Dr. Lai received a B.S. degree in mathematics and chemistry from Lyon College in 1995, and a Ph.D. degree in chemistry from the University of North Carolina at Chapel Hill in 1999. She subsequently joined IBM, where she is working on gate dielectric reliability.

David L. Harmon *IBM Microelectronics Division, 1000 River Road, Essex Junction, Vermont 05452 (dlharmon@us.ibm.com).* Mr. Harmon received a B.S. degree in 1977 and an M.S. degree in 1979, both in electrical engineering, from the University of Maine. Since joining

IBM in 1979, he has worked in the areas of plasma process development, ASIC applications, and reliability engineering. He is an author or co-author of ten patents and more than 30 technical publications. Mr. Harmon is currently a Senior Engineer with the IBM Microelectronics Division and is involved in the study of integrated circuit reliability mechanisms.