Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Microelectronics Reliability 45 (2005) 427–436
MICROELECTRONICS RELIABILITY
www.elsevier.com/locate/microrel

Introductory Invited Paper

Gate dielectric breakdown in the time-scale of ESD events

Bonnie E. Weir [a,*], Che-Choi Leung [b], Paul J. Silverman [a], Muhammad A. Alam [c]

[a] Agere Systems, 555 Union Boulevard, Allentown, PA 18109, United States
[b] Agere Systems, 1110 American Parkway N.E., Allentown, PA 18109, United States
[c] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, United States

Received 3 November 2004
Available online 12 January 2005

Abstract

Transmission line pulse (TLP) measurements are used to demonstrate that oxynitride breakdown projections from DC measurements using conventional area and voltage-scaling techniques can be extended to the nanosecond time-scale. ESD protection systems can thus be designed to prevent dielectric breakdown. Important concepts in gate dielectric breakdown such as the anode–hole injection model and area and statistical effects are discussed and applied to the nanosecond regime.
© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

The challenge of electrostatic discharge (ESD) design is that as transistor scaling continues and operating voltages are lowered, the interface to the outside world and therefore the ESD specifications remain the same. Moreover, ESD protection has aimed at avoiding the first breakdown voltage for snapback of a transistor. However as has been highlighted by Duvvury and Amerasekera [1] this snapback voltage and the median breakdown voltage of the gate dielectric are converging (Fig. 1), making it difficult to ensure the robustness of gate dielectrics in an ESD event. This is particularly true for 1 V I/O's in high-speed, high-performance applications, where transistor gates may be directly connected to an external pin or NFETs may be used as compact decoupling capacitors between VDD and VSS, exposing the thin gate dielectric to ESD pulses which are characterized by very high-current and very short duration. During an ESD pulse, the ESD protection and interconnects can both contribute voltage drops, and their sum must not be higher than the voltage which a dielectric can withstand.

A vast knowledge base exists for oxide breakdown in the long time-scale, and a few publications have addressed short time-scales [1–8]. For the most part, studies which described breakdowns at short times have used methodology developed for thicker oxides. In the case of thinner oxides or oxynitrides at voltages below ~5 V, the voltage acceleration is no longer $1/E$, and small differences in thickness have a significant effect on the dielectric breakdown voltage. The theory which has emerged from thin oxide studies has not been applied before by the ESD community and statistical and area-dependent effects have generally been neglected. Therefore, the purpose of this paper is to study the relevance of long time-scale TDDB data in predicting the response to short time-scale ESD events, especially for sub-2 nm dielectrics in both NFETs and PFETs. A summary of important concepts for dielectric breakdown will be given and

* Corresponding author. Tel.: +1 610 712 2042.
E-mail address: bew@agere.com (B.E. Weir).

0026-2714/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.microrel.2004.12.004

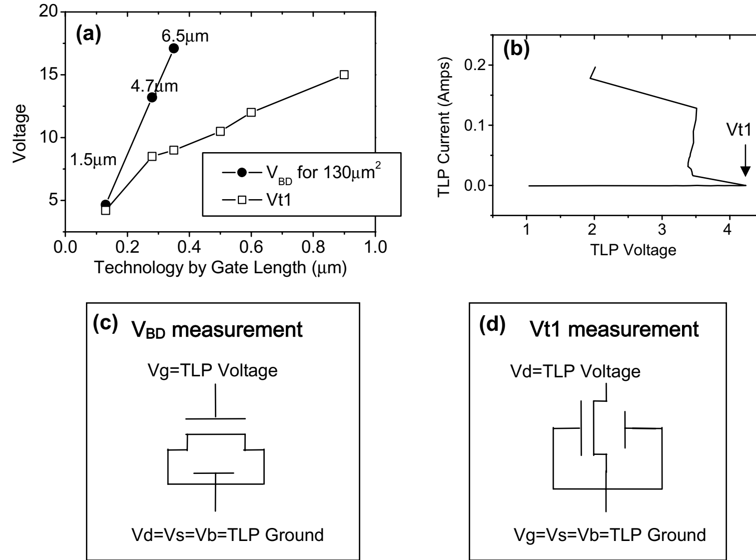

Fig. 1. (a) First breakdown voltage (V_{t1}) for NFET snapback and dielectric breakdown voltage V_{BD} vs. gate length for several technology generations of NFETs. Both voltages were measured using a TLP system with 3 ns risetime and 100 ns pulses. V_{BD} measurements are also shown in Fig. 3. A measurement of V_{t1} is shown in (b) for a 1.5 nm oxynitride in a 0.13×2000 μm device. V_{t1} is sensitive to underlying doping as well as to channel length; (c) measurement configuration for V_{BD} and (d) for V_{t1}.

experimental results from studying 1.5–1.9 nm dielectric breakdown in the nanosecond regime will be presented. It will be shown that many of the breakdown concepts developed for the long time-scale continue to apply to the nanosecond regime. Methods for projection to thinner dielectrics (1.2 nm) will be discussed.

2. Background

Several important concepts will be used in the discussion of dielectric breakdown in the nanosecond regime. These concepts will be reviewed here before the new experimental results are presented.

2.1. Voltage acceleration

The anode–hole injection model and the validity of various approximations is treated extensively in Ref. [9]. A description of the model for $V < 9$ V will be given here. A diagram to illustrate the process is shown in Fig. 2a. The picture for thin oxides is that as electrons tunnel across the oxide, some of those electrons create electron–hole pairs. A portion of those generated holes are injected from the anode back into the oxide and can tunnel across (J_h). A fraction of those tunneling holes generate traps in the oxide. When the oxide is bridged by overlapping traps, we call it breakdown. The following equation describes the relationship between time to breakdown (T_{BD}) and electric field in the anode–hole injection model.

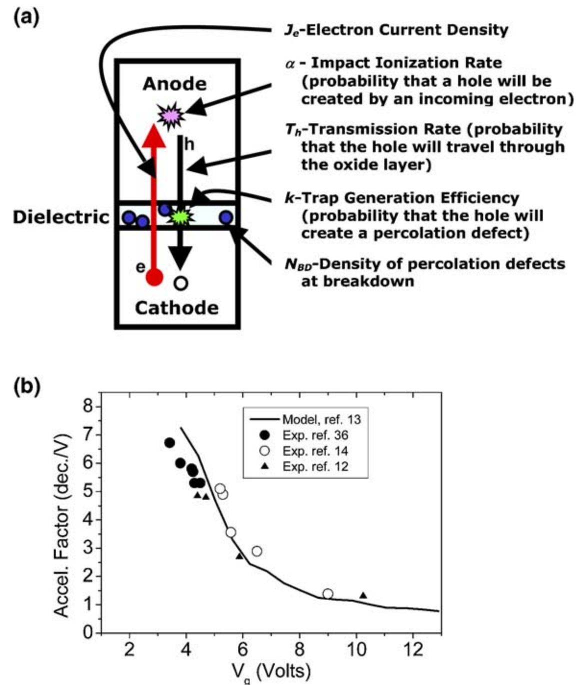

Fig. 2. (a) Illustration of the anode–hole injection model showing the various terms in Eq. (1). (b) Theoretical calculation of voltage acceleration factor for time to breakdown from Ref. [13] as well as experimental results from several groups [10,12,14].

$$T_{BD} = \left(\frac{N_{BD}}{k}\right)^{\frac{1}{m}} \frac{1}{J_h} \quad (1)$$

where N_{BD} is the critical trap density, a quantity that varies statistically and can be predicted from the percolation model which is explained in the following section. The statistical variation is independent of field. The lattice–hole interaction coefficient, k, indicates how difficult it is for a tunneling hole to generate a trap. Since the rate of trap generation may vary with stress time, the nonlinearity coefficient, m, is included. The number of tunneling holes can be described by $J_h = J_e \alpha T_h$, where J_e is the density of tunneling electrons, α is the probability of electron–hole pairs being generated, and T_h is the transmission probability of holes. When the electron tunneling is Fowler–Nordheim, $J_e = A\exp(-B/E)$ where A and B are constants and E is the electric field. So $\log(T_{BD}) \sim 1/E$ is valid at voltages where essentially all electrons which tunnel from the cathode generate electron–hole pairs in the anode ($\alpha = 1$ or has no field-dependence) and all of those generated holes tunnel back across the oxide ($T_h = 1$ or has no field-dependence). In fact, in the range where this study has been performed, around 5 V, although the electron tunneling continues to be Fowler–Nordheim, both α and T_h, depend on E. The time to breakdown equation is dominated by $\alpha = M\exp(DV)$ where M and D are constants, and therefore $\log(T_{BD}) \sim V$ as has been reported by many different groups [10–12]. Fig. 2b shows the voltage acceleration as a function of voltage plotted by Alam et al. [13] based on a theoretical evaluation of all of the terms considered above. Measurements from several groups are also included in Fig. 2b [10,12,14]. A further complication as the voltage is lowered below 4 V is that α begins to depend even more steeply on voltage so that $T_{BD} \sim V^{-44}$ as has been reported by Wu et al. [15]. This becomes important for projections to operating voltage, but not for nano–second time–scale breakdown for oxynitrides down to 1.2 nm.

2.2. Polarity dependence

We include the ESD response for both NFETS and PFETS because PMOS reliability for sub-2 nm oxynitrides is a particularly important topic [16]. Although PFETs are more robust in the 5–7 V range, the time to breakdown, at operating voltage, of PFETs is shorter than that of NFETs because minority ionization can occur in PFETs. This process creates an energetic hole, and although very rare, becomes significant at low voltages [17].

The measurements and validation discussed above are based on DC stress, unlike the ESD pulse of very short duration. Therefore, the fundamental question is whether or not the DC formulation is relevant for ESD breakdown. In this paper, we directly measure breakdown in the ns regime. We confirm that the same acceleration rules can be used both in the DC and ns regimes.

2.3. Percolation model

For more than a decade, oxide breakdown results have been interpreted based on the percolation model. Many different groups have validated this model [18–22] and the data here will be interpreted based on this model. Some of the implications of the model will also be explored. The percolation model is as follows: Traps form randomly in space throughout the dielectric. As more and more traps are formed, the traps begin to overlap. When the overlapping traps bridge the dielectric, a high-current path (percolation path) forms between the two electrodes and breakdown occurs (Fig. 3). The percolation model accurately predicts the statistical variation observed in oxide breakdown data as well as the area-dependence [18]. The statistical variation in breakdown time or breakdown voltage simply means that if a number of identical dielectrics are stressed under identical conditions, statistically, percolation paths will form earlier in some than in others. Because the breakdowns follow a statistical pattern, the time for a small fraction of devices to exhibit breakdown can be predicted from the median time to breakdown and the Weibull shape factor [20]. In addition, increasing the sample area is similar to increasing the statistical pool and therefore results in a shorter breakdown time. The observed area-dependence of the breakdown time is correctly predicted by the percolation model [18].

2.4. Importance of statistics

Although most papers in the ESD field have focused on determining the median breakdown voltage of a dielectric, it is extremely important to take into account the statistical variation of breakdown and the area-dependence. As shown in Fig. 4 from Wu et al. [18], the Weibull shape factor decreases as oxide thickness

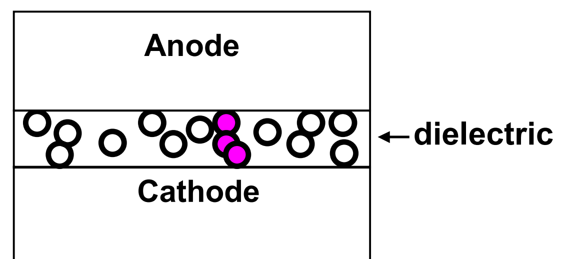

Fig. 3. Traps are depicted as circles within the dielectric. The traps which form a percolation path across the dielectric are shaded. The path may contain more than three traps depending on the arrangement.

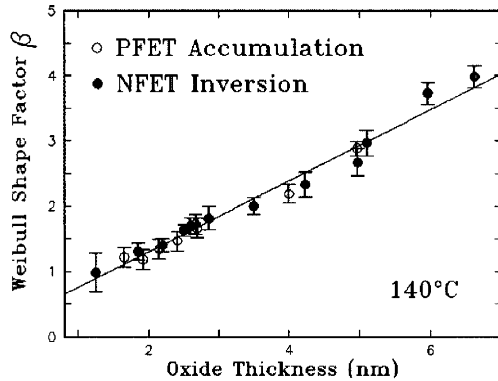

Fig. 4. Weibull shape factor vs. oxide thickness from Wu et al. [18] illustrating the concept that as the dielectric becomes thinner, the statistical spread in time to breakdown increases.

is reduced. The breakdown voltage is related to the Weibull shape factor, β, by the following equation:

$$\Delta V_{BD} = \frac{1}{\text{Volt. Acc.}} \times \frac{1}{\beta} \log \left(\frac{\ln(1 - f_1)}{\ln(1 - f_2)} \right) \qquad (2)$$

The difference in breakdown voltage for a 5 nm oxide, as shown in Fig. 5, is almost 2 V. Although this effect is somewhat smaller for thinner oxides, since the voltage acceleration factor increases at these voltages, it still remains a >10% effect. While V_{t1} was much lower than V_{BD} (e.g. 0.3 μm technology node $V_{BD} - V_{t1} \sim 7$ V) statistical effects which changed V_{BD} by ~2 V could be ignored. However, in the regime where V_{t1} is close to V_{BD}, these statistical effects can make the difference between a product passing the ESD specification or not. These statistical effects are taken into account in our study and included in the projection to 1.2 nm.

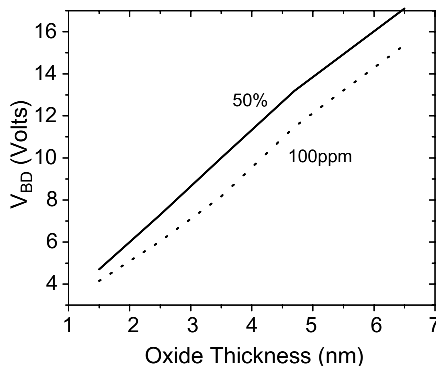

Fig. 5. V_{BD} vs. dielectric thickness for 100 ns stress on NFETS. The spread between the median value and a standard industy specification is shown. Measurements shown in Fig. 1a are used for the 50% line, and the conversion to 100 ppm use the Weibull shape factor values from Fig. 4.

2.5. Area-dependence

As described above, the breakdown time or breakdown voltage also depends on the area of the dielectric under test. The difference in breakdown voltage is shown in Eq. (3).

$$\Delta V_{BD} = \frac{1}{\text{Volt. Acc.}} \times \frac{1}{\beta} \log \left(\frac{A_1}{A_2} \right) \qquad (3)$$

In oxide reliability studies, we use the area-dependence to relate the breakdown measured on a test structure, to the area of the entire chip. This is valid, because the stress voltage being considered is simply the operating voltage which will be applied all over the chip. In an ESD event, the voltages are more local, and the affected area may be similar to the tester area. However, there may be cases where the gate dielectric area affected by an ESD event may be more than an order of magnitude larger or smaller than the tester area and therefore the area-dependence is included in this work.

2.6. Soft breakdown

One topic which has been extensively studied recently, in connection with thin oxides, is soft breakdown [23–27]. This is when a percolation path is completed, but the thermal effects associated with traditional dielectric breakdown do not take place and the transistor remains functional. Soft breakdown occurs at low voltages, or when the current flowing through the oxide is limited by instrumental compliance or by a series resistance [23–25]. After soft breakdown, the current–voltage characteristic shows a power-law relationship ($I = V^{\delta}$ with $\delta = 4$–7), whereas, hard breakdown results in an ohmic characteristic ($\delta \sim 1$). Soft breakdown is not relevant for ESD events which cause breakdown in 1.2–1.9 nm oxynitrides, due to the breakdown voltage. Hard breakdown has been observed at voltages as low as 2 V [28]. Soft breakdown can occur at voltages above 4 V if current is limited [25], but high currents are certainly available in ESD events. Therefore, the conditions of low voltage and limited current do not apply to ESD events, and soft breakdown is not generally relevant to the topic.

2.7. Previous work in the short time-scale

A few papers, without considering ESD events, have included a discussion of the oxide breakdown transient, Ridley [3] calculated that breakdown should occur in approximately 10 ns. Experimentally, Lombardo et al. [29] recently stated that above 4 V, the breakdown transient is greater than 10^5 A/s. These studies indicate that a breakdown event could be completed within the timescale of an ESD event. Below, a number of papers which consider oxide breakdown directly in the context of an ESD event are discussed.

Cheung [6] has demonstrated that the gate oxide degradation mechanism remains unchanged in the short time-scale. His work used an indirect measure of oxide damage developed in order to study plasma damage [6]. He obtained voltage acceleration factors matching the universal voltage acceleration curve in Fig. 2b as he compared damage due to nanosecond stresses and DC stresses. His paper is the first, to our knowledge, which includes an explanation of statistical and area-dependent effects in the context of oxide breakdown in an ESD event.

Matsuzawa et al. [7] studied a 3 nm oxide and showed that the breakdown mechanism was the same over the range of 10^{-4} to 10^{-7}s by using repeated TL pulses. They used an equation similar to our Eq. (1) to simulate the time to breakdown as a function of voltage and find good agreement for the 3 nm oxide. The lack of agreement between their data and the simulation for the 6 nm oxide remains a topic for further investigation. They also include a helpful discussion on the edge effects in a transistor.

Salman et al. [30] have studied the oxide robustness in an ESD event. They include a figure similar to our Fig. 1a) which compares oxide breakdown voltage with V_{t1} and holding voltage as a function of technology generation. They show an example of a circuit with a conventional snapback ESD protection system where the failure mode is oxide rupture although the median oxide breakdown voltage was slightly above V_{t1} for that technology (0.1 μm). For longer gate length protection structures, the differential resistance causes the holding voltage to exceed the oxide breakdown voltage at high ESD current. Consequently, oxide breakdown occurs in the gate drain overlap region of the protection NFET.

Wu and Rosenbaum [2] concluded that for a 2.2 nm oxide, the $1/E$ dependence holds over the range of 10^4 to 10^{-9} s. To our knowledge, this study covers the widest range of time to breakdown reported to date. They also show that self-heating effects are not important in understanding oxide breakdowns at these voltages and time-scales. PMOS devices in accumulation are studied, because accumulation stress has a lower V_{BD} for PMOS. In the voltage range they studied, the time to breakdown is expected to exhibit an approximately $1/E$ dependence. However, as even lower voltages are probed, this $1/E$ dependence, which comes from a simplification of the anode–hole injection model, no longer holds [9] as described above.

Additionally, they state that with sufficient guard-banding, statistical effects are not a point of concern [2]. Although this is certainly true, it becomes increasingly expensive in real estate and sometimes impossible to build large voltage margins into the ESD protection system. For example, if the median breakdown time of a given oxide is 4 V, in order to provide sufficient guard-banding, without calculating statistical effects, the ESD protection system would have to be designed such that the voltage would never rise above 3 V. There is a significant advantage in doing the simple calculations outlined here to determine the area and statistical effects in order to see that the protection system must be designed to avoid 3.4 V, for example.

Our experiments are described in the following sections. We have focused on the aspects of dielectric breakdown most relevant to ESD design for current technologies. Both NFETs and PFETs with 1.5 nm oxynitrides are tested in inversion. Methodology for projection to even thinner dielectrics is discussed.

3. Experiment

3.1. Transmission line pulse measurements

Transmission line pulse (TLP) testing has been used to study the effect of an electrostatic discharge [31]. Our TLP system applies 100 ns pulses (3 ns rise time) to a device, while measuring the voltage across and current through it. The robustness of gate dielectrics on a very short time-scale was investigated by subjecting them to incrementally increasing voltage pulses, and making DC leakage measurements between pulses, as depicted in Fig. 6. Testers of 0.13 μm NFETs and PFETs with oxynitride dielectrics were designed specifically for this experiment. One contact pad was wired to the gate and another contact pad to the source, drain and tub to comply with the 2-probe TLP configuration. Series resistance was minimized by using multiple 0.13×10 μm devices so that metal contact to poly-silicon was never more than 5 μm away. Testing in inversion also ensured that the resistance associated with a

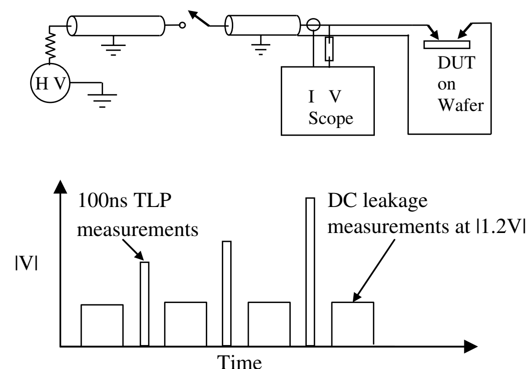

Fig. 6. Wafer-level transmission line pulse system used in these experiments. The 50 Ω environment is preserved down to within a few mm of the device using triaxial probes. The TLP system applies a 100 ns pulse during which voltage and current are measured. Between TLP pulses, a DC leakage measurement is performed.

substrate or n-well contact was not included. In a TLP system, the voltage and current levels depend on the relative impedance of the transmission line (TL) with respect to the device under test (DUT): $I_{DUT} = V_C/(Z_{DUT} + Z_{TL})$ where V_C is the charging voltage. If a gate dielectric has a resistance $Z_{DUT} = 10\ \text{k}\Omega$, with $V_C = 4.7$ V and $Z_{TL} = 50\ \Omega$, the total resistance is dominated by the resistance of the DUT, so the stress is equivalent to a constant-voltage stress. V_{DUT} was measured to be within 0.1 V of V_C as further confirmation of minimal series resistance. We can therefore compare the results for the first dielectric breakdown under TLP stress to DC constant-voltage stress [32].

3.2. High resolution measurements of the breakdown event

Dielectric breakdown was directly observable in the TLP waveforms as shown in Fig. 7. The finer time-resolution used here confirms the finding of Lombardo et al. [29]. The breakdown event can occur in less than 2 ns and has a transient as high as 3.5×10^6 A/s. The slope of the transient is similar for both 5 V and −6 V, although transients at lower voltages may be significantly less steep [29]. If the breakdown transient were more than 100 ns, it is possible that the breakdown process would not have been completed within one stress cycle, resulting in significant differences between breakdown measured in the DC regime and that in the nanosecond regime. With this time-scale established for breakdown, we see no reason why the damage caused by a pulse which lasts 100 ns would not be similar in character to the damage caused by a longer stress. However, since some of the measured transients lasted several nanoseconds, if we had been measuring breakdown due to a 1 ns pulse, DC extrapolation may not have been possible. Since some ESD transients only last a few nanoseconds, further study is required, before the methods used here are extrapolated to the 1–2 ns regime. These data indicate that DC extrapolation techniques can be used at least for pulse widths of 10 ns or greater.

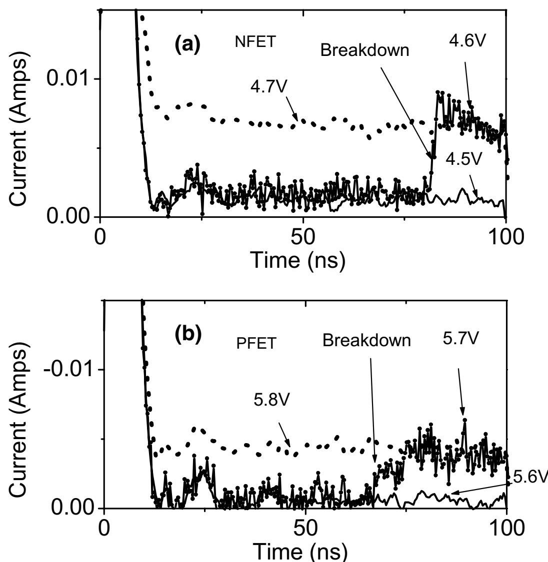

Fig. 7. TLP waveforms show dielectric breakdown with high-time resolution in (a) an NFET and (b) a PFET. The voltage steps before and after the step in which the first dielectric breakdown occurred are plotted. The observed increase in current (labeled "breakdown") is accompanied by a decrease in voltage. After such an event, the DC leakage at |1.2 V| increases to the $\pm 1 \times 10^{-4}$ A range. For pulses which showed no increase in current, a 5-point averaging routine was used to smooth the data to improve the clarity of these figures. The initially high-current value is an artifact of the high-frequency measurement and occurs because at the measurement location, the incident and reflected pulses do not overlap perfectly. Although a relatively fast breakdown transient is shown for the NFET and a slower transient shown for the PFET, no difference in average transient speed was detected between the two types of device. The breakdown transients from both NFETs and PFETs were in the 2–10 ns range.

3.3. Ramped voltage measurements

We also assume that the bulk of the damage occurs in the final step of the ramp. For example, if a device undergoes hard breakdown within 100 ns at 4.8 V, using the voltage acceleration factor of 5.8 dec./V, we find that it would last 380 ns at the previous pulse of 4.7 V. Therefore, the damage at 4.7 V is only about 1/4 of that at 4.8 V, at 4.6 V it is only $1/4n$ of that at 4.8 V (where n = number of steps between 4.6 and 4.8 V), and we can consider the main damage to occur in the final voltage step. To confirm the validity of this assumption, we have made measurements by repeatedly pulsing at the same voltage. The time to breakdown obtained from these measurements at 4.6 V and 4.7 V falls on the same voltage acceleration line as the ramped-voltage measurements as shown in Fig. 9. Although at higher voltages this assumption may not have held true, and the stress involved at each step of the ramp needed to be calculated in order to compare ramped voltage measurements and constant voltage measurements [33], it is sufficient to consider the final step of the ramp in this regime where the voltage acceleration is >5.8 dec./V.

4. Results

4.1. DC vs. nanosecond measurements—voltage acceleration

Fig. 8 shows the data for 0.13 μm NFETs and PFETs with 1.5 nm oxynitride gates. All dielectric breakdowns measured by TLP were hard (δ = 1–2.5 where $I = V^{\delta}$).

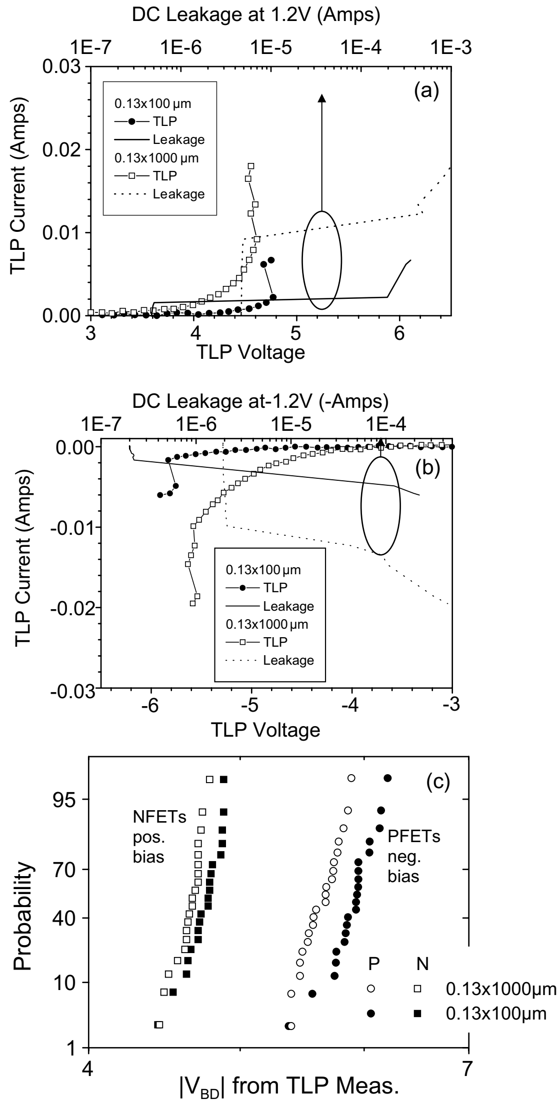

Fig. 8. TLP data showing 1.5 nm oxynitride breakdown (a) TLP current vs. TLP voltage for the two NFET areas (pos. voltage stress). DC leakage is shown on the top axis and is plotted against TLP current; (b) similar plot for PFETs (neg. voltage stress); (c) lognormal plot of $|V_{BD}|$, the point at which at least a 10x increase in DC leakage was observed. The 50% probability values are listed in Table 1. Two areas were tested for both NFETs and PFETs.

As shown, the leakage at 1.2 V (plotted on the top axis vs. TLP current) increases by several orders of magnitude at the point of breakdown. Although in many cases several dielectric breakdowns were observed, the voltage where first dielectric breakdown occurred was considered V_{BD}. We compared V_{BD} to our expectations based on DC measurements at lower voltages [34]. The voltage acceleration data are shown in Fig. 9 and agree with published values [10,35–37]. The voltage acceleration

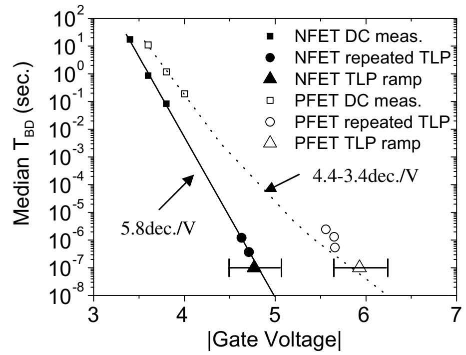

Fig. 9. DC constant voltage measurements [34] of time to breakdown for 1.5 nm oxynitrides are fit to determine the voltage acceleration. TLP measurements on $0.13 \times 100\ \mu m^2$ devices are plotted just below 100 ns. The overlap of the TLP measurement with the projection from DC measurements confirms that even nanosecond-range breakdowns can be predicted using standard projection techniques. In addition, median breakdown times from repeated 100 ns pulses at 4.6 and 4.7 V for NFETs and 5.65 and 5.55 V for PFETs are plotted, demonstrating that with this steep voltage acceleration, the bulk of the damage occurs in the final step of the ramp.

lines are extrapolated to the 100 ns range. The V_{BD} values from Fig. 8 are plotted with error bars and overlap the DC projections. Also, repeated TLP pulses at 4.6 and 4.7 V for NMOS and 5.5–5.7 V for PMOS were applied to determine the time to breakdown at these voltages. The results fall on the same line as the voltage acceleration determined in the longer time-scale. In addition, V_{BD} determined from the ramped voltage measurements is in agreement with these points if we consider the bulk of the stress to have occurred in the final 100 ns step of the ramp and assign a time to breakdown of 100 ns. The PFET has a higher V_{BD} than the NFET in this voltage range due to the lower tunneling current at a given $|V|$. As shown, the voltage acceleration rules used for the DC case are applicable to the nanosecond measurements.

4.2. Area-dependence for nanosecond measurements

The area-dependence is analyzed in Table 1. Using a Weibull shape factor of 1.2 obtained from DC measurements for this technology and in agreement with Fig. 4, we calculated the V_{BD} we would expect for a device with 10 times larger area. For the NFET the calculated values are very close to the measured values. For the PFET, the calculated difference between the two areas, 0.25 V, is similar to the measured difference of 0.26 V. Therefore, area scaling methods have been shown to apply to the nanosecond regime. With the use of thin oxides where gate oxide rupture must be carefully avoided during an ESD event, accurate calculation of the breakdown

Table 1
Comparison of TLP measurements with calculations based on DC measurements for both N and PFETs

	TLP exp. from Fig. 8		DC Time to Breakdown data		Predicted from DC	
	1	2	3	4	5	6
Device	Median V_{BD} from TLP for 0.13×100 μm^2 (Volts)	Meas. V_{BD} from TLP for 0.13×1000 μm^2 (Volts)	Voltage acceleration (decades/V)	t_{BD} from DC meas (sec.)	Calc. V_{BD} for 0.13×100 μm^2 (Volts)	Calc. V_{BD} for 0.13×1000 μm^2 (Volts)
NFET	4.77 ± 0.3	4.66 ± 0.3	5.8 ± 0.7 near 3.6 V	0.084 at $V1 = 3.8$ V	$V2 = 4.8 \pm 0.2$	$V3 = 4.7 \pm 0.2$
PFET	-5.93 ± 0.3	-5.67 ± 0.3	4.4 ± 0.7 near -4 V to 3.4 ± 0.4 near -5.5 V	0.19 at $V1 = -4.0$ V	$V2 = -5.8 \pm 0.3$	$V3 = -5.55 \pm 0.3$

Columns 1 and 2 are obtained from Fig. 8. Columns 3 and 4 are obtained from Fig. 9, although the PFET includes the curvature at higher voltages [13]. Columns 5 and 6 are calculated from the other columns using the following equations: (Column 5: using Volt. Acc. from Column 3, t_{BD1} and $V1$ from Column 4, and $t_{BD2} \sim 25$–100 ns, $V2 = V1 + (1/\text{Volt. Acc.}) \times \log(t_{BD1}/t_{BD2})$) and (Column 6: using Volt. Acc. from Column 3 and $\beta = 1.2$, $V3 = V2 - (1/\text{Volt.Acc.}) \times (1/\beta)\log(A3/A2)$).

voltage, including area-dependence, is extremely important. If area considerations are neglected, as has been done in the past, 0.1–0.5 V errors can be made in V_{BD}.

4.3. Thickness scaling

Dielectrics in NFETs ranging in thickness from 1.5 to 1.9 nm were tested in the DC regime and the nanosecond regime as shown in Fig. 10. In the DC regime, all were tested at 3.8 V and a thickness acceleration of 6.3 dec./nm was found. The physical thickness for the 1.7 nm and 1.9 nm dielectrics was determined by TEM analysis. The physical thickness of the 1.5 nm dielectric was estimated based on the measured differences in electrical thickness. In the ns regime, a ramped-voltage measurement was performed, and the slope of the median breakdown voltage vs. dielectric thickness was found to be 1.9 V/nm. Combining the two plots, we obtain a voltage acceleration of 3.3 dec./V, which is in agreement with the universal voltage acceleration curve shown in Fig. 2b. These data again demonstrate that the breakdown physics remains the same in the ns regime and that the parameters obtained from DC measurements can be used to extrapolate to the time-scale of interest for ESD events.

4.4. Series resistance

One of the common techniques used to protect a sensitive device from snapback-induced failure is to add extra resistance between the ESD protection device and the device being protected. The resistor, R, is typically 10–200 Ω and it limits the current flowing into the device being protected. For example, if the ESD protection device has a resistance of 1 Ω during an ESD discharge, and the resistor R is 100 Ω, then only 1/100th of the total ESD discharge current could flow into the device being protected. For a 2 kV HBM discharge, the maximum current the device being protected will see is <13.3 mA. If the device failure current, I_{t2}, is larger than 13.3 mA then the device will be safe from ESD damage. A typical I_{t2} value for a 0.13 μm NMOS is ~1 mA/μm of transistor channel width, so this is a very effective method to enhance the ESD protection. However, the resistance of a gate dielectric is in the kΩ range, and it requires much less power to damage. Therefore, it will take a much larger R to limit the ESD current and power flowing into the device being protected to a safe level. Fig. 11 shows the median breakdown voltage as a function of the additional series resistance. In the case of 1.5 nm oxynitride breakdown, the traditionally used values of 100 Ω series resistance do not substantially alter the breakdown voltage. The 1 kΩ resistor does increase the breakdown voltage to 6.3 V. For all of the 0.13×100 μm area dielectrics tested without intentional series resistance, the average resistance value obtained from the TLP current and voltage immediately prior to breakdown was 3.2 kΩ. An additional 1 kΩ is therefore predicted to break down at a voltage given by $4.77 \times 4.2/3.2$ dividing the voltage between the dielectric and the additional resistor. The result of 6.26 V is in close agreement with the experimental result of 6.33 V.

5. Implications

5.1. Projection to thinner dielectrics

Using all of this information, a projection can be made to a thinner dielectric. Although, gate dielectric reliability of PFETs is worse than that of NFETs below

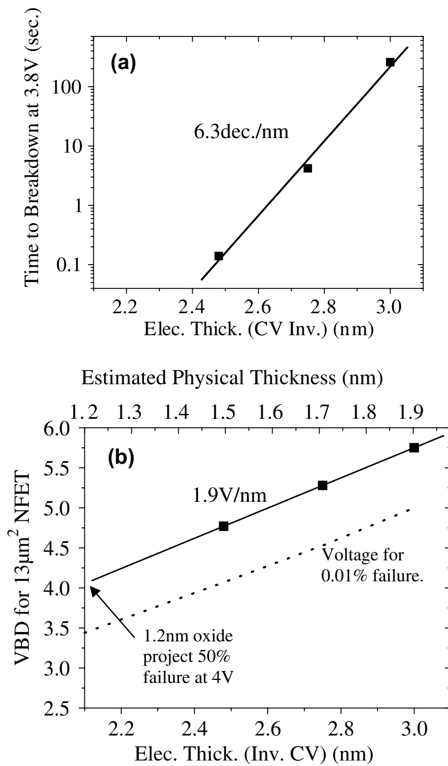

Fig. 10. (a) DC measurements of median time to breakdown vs. electrical thickness for NFETs with oxynitride gate dielectrics. All measurements were at 3.8 V and were either measured on or projected to 13 μm^2 device area; (b) TLP measurements of breakdown voltage for 13 μm^2 device area vs. electrical thickness. The top axis shows the physical oxide thickness corresponding to the electrical thickness measured by *CV* in inversion. The solid squares represent the median breakdown voltage and the dotted line represents the calculated voltage for 0.01% failure. TEM measurements are compared to electrical measurements for the 1.7 and 1.9 nm oxynitrides and the relationship is used to determine the physical thickness for the 1.5 nm structure.

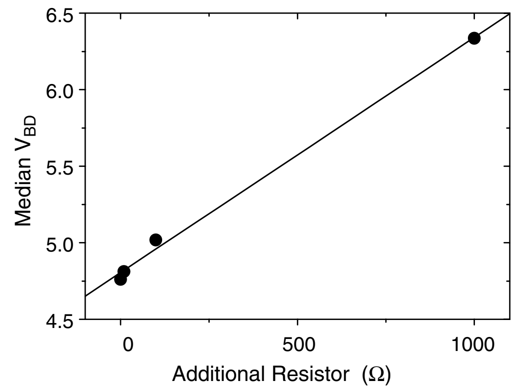

Fig. 11. TLP testing results from testers which included a series resistor in the layout. The median of >10 devices is shown for each resistor value. 0.13 × 100 μm NFETs with 1.5 nm oxynitrides were stressed with 100 ns TL pulses and 0.1 V steps.

~3 V, NFET gate dielectrics are clearly more sensitive, in the region of interest for ESD events, as shown in Fig. 9. Therefore our projections will be for NFETS. Assuming that the thickness acceleration obtained in the 1.5–1.9 nm regime is valid down to 1.2 nm, as shown in Fig. 10, the median breakdown voltage for a 100 ns stress is projected to be 4 V for a 13 μm^2 device. The dotted line in the figure shows the voltage at which we would predict 0.01% failure. Eq. (2) is used to obtain this line. The Weibull slope from Fig. 4 and the voltage acceleration from Fig. 2b are used. The Weibull slope becomes shallower for thinner dielectrics, but is offset by the steeper voltage acceleration at lower voltages. The result of the calculation is that the voltage across the oxide must be kept below 3.4 V for a 13 μm^2 area in order to keep oxide breakdown below the 100 ppm level.

6. Conclusions

We have reviewed important concepts for oxide breakdown, such as the anode–hole injection model, the area and statistical dependence of breakdown and soft breakdown. We have shown that projections of dielectric breakdown from DC measurements are in agreement with nanosecond measurements of hard breakdown and that the breakdown transient can take less than 2 ns. With the knowledge that breakdown in the nanosecond regime follows the well-understood voltage acceleration, area-dependence and thickness dependence, we can expect that the Weibull statistics are obeyed as well and dielectric breakdown can be fully included in ESD design using standard extrapolation techniques.

Acknowledgments

Helpful suggestions and contributions from Robert Ashton, Evan Grand, Tanya Nigam, Charvaka Duvvury, Bill Hunter, J. DeLucca, and K. Hooghan are gratefully acknowledged.

References

[1] Duvvury C, Amerasekera A. Semicond Sci Technol 1996;11:833.
[2] Wu J, Rosenbaum E. IEEE Trans Electron Dev 2004;51: 1528–32.
[3] Ridley BK. J Appl Phys 1975;46:998.

[4] Chen Y, Suehle JS, Shen C-C, Bernstein J, Messick C, Chaparala P. International Reliability Physics Symposium Proceedings, 1998. p. 87–91.
[5] Wang B, Suehle JS, Vogel EM, Bernstein JB. IEEE Electron Dev Lett 2001;22:224.
[6] Cheung KP. Microelectron Reliab 2001;41:745.
[7] Matsuzawa K, Satake H, Sutou C, Kawashima H. SISPAD Proceedings, 2003. p. 129.
[8] Wu J, Juliano P, Rosenbaum E. Electrical Overstress/Electrostatic Discharge Symposium Proceedings, 2000. p. 287.
[9] Alam M, Weir B, Bude J, Silverman P, Ghetti A. Microelectron Eng 2001;59:137–47.
[10] Weir BE, Alam MA, Bude JD, Silverman PJ, Ghetti A, Baumann F, Diodato P, Monroe D, Sorsch T, Timp G, Ma Y, Brown MM, Hamad A, Hwang D, Mason P. Semicond Sci Technol 2000;15:455–61.
[11] Schlund BJ, Suehle J, Messick C, Chaparala P. Microelectron Reliab 1996;36:1655–8.
[12] Hu C, Lu Q. International Reliability Physics Symposium Proceedings, 1999. p. 47–51.
[13] Alam MA, Bude J, Ghetti A. International Reliability Physics Symposium Proceedings, 2000. p. 21–6.
[14] Nigam T. Ph.D. Thesis. Katholieke Universiteit Lueuven, May 1999.
[15] Wu EY, Vayshenker A, Nowak E, Sune J, Vollertsen RP, Lai W, Harmon D. IEEE Trans Electron Dev 2002;49:2244–53.
[16] Weir BE, Silverman PJ, Alam MA, Ma Y. Proc Electrochem Soc 2002:465.
[17] Bude JD, Weir BE, Silverman PJ. IEEE Electron Devices Meeting Proceedings, 1998. pp. 179–82.
[18] Wu EY, Sune J, Lai W. IEEE Trans Electron Dev 2002;49:2141–50.
[19] Degraeve R, Groeseneken G, Bellens R, Ogier JL, Depas M, Roussel PI, Maes HE. IEEE Trans Electron Dev 1998;45:904–11.
[20] Alam MA, Weir BE, Silverman P. IEEE Circuits Device 2002;18:42–8.
[21] Stathis JH. J Appl Phys 1999;86:5757–66.
[22] Dumin DJ, Maddux JR, Scott RS, Subramoniam R. IEEE Trans Electron Dev 1994;41:1570–80.
[23] Alam MA, Weir BE, Silverman PJ. IEEE Trans Electron Dev 2002;49:232–8.
[24] Alam MA, Weir BE, Silverman PJ. IEEE Trans Electron Dev 2002;49:239–46.
[25] Linder BP, Stathis JH, Wachnik RA, Wu E, Cohen SA, Ray A, Vayshenker A. VLSI Technology Symp Tech Dig 2000:214–5.
[26] Kaczer B, De Keersgieter A, Mahmood S, Degraeve R, Groeseneken G. International Reliability Physics Symposium Proceedings, 2004. p. 79–83.
[27] Sune J, Wu EY, Lai WL. IEEE Trans Electron Dev 2004;51:1584–92.
[28] Wu E, Sune J, Linder B, Stathis J, Lai W. Electron Devices Meeting, Tech. Dig. 2003. p. 38.1.1–4.
[29] Lombardo S, Stathis JH, Linder BP. Phys Rev Lett 2003;90:167601.
[30] Salman A, Gauthier R, Wu E, Riess P, Putnam C, Muhammad M, Woo M, Ioannou D. International Reliability Physics Symposium Proceedings, 2002. p. 170–4.
[31] Barth J, Verhaege K, Henry LG, Richner J. Electrical Overstress/electrostatic Discharge Symposium Proceedings, 2000. p. 85–96.
[32] During and after first dielectric breakdown, when the DUT resistance decreases, constant voltage is not maintained.
[33] Berman A. International Reliability Physics Symposium Proceedings, 1981. p. 204.
[34] For Fig. 9, PFET DC measurements in inversion at negative voltage used 0.13×100 μm devices while NFET measurements at positive voltage were performed on 0.13×20 μm structures. Area-scaling was performed using $(A1/A2)^{(1/\beta)}$ for the NFET with $\beta = 1.2$.
[35] For the projections for the PFET, which spans 2 V rather than 1 V, we have taken into account the curvature of the voltage acceleration. Similar values can be found in.
[36] Stathis JH. International Reliability Physics Symposium Proceedings, 2001. p. 132.
[37] Nicollian PE, Hunter WR, Hu JC. International Reliability Physics Symposium Proceedings, 2000. p. 7.