Gate Oxide Evaluation under Very Fast Transmission Line Pulse (VFTLP) CDM-Type Stress

Slavica Malobabić, David F. Ellis, Javier A. Salcedo, Yuanzhong (Paul) Zhou, Jean-Jacques Hajjar, and Jun J. Liou
(a) School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, Florida, USA
(b) Analog Devices, Wilmington, Massachusetts, USA

Abstract- Gate oxide breakdown is analyzed under very fast transmission line pulsed (VFTLP) stress, using different pulse -rise times and -widths. The switching of oxide behavior pre- and post- breakdown occurs in tenths of a nanosecond and it shows reproducible voltage and current characteristics. The total stress and time-dependent-dielectric-breakdown (TDDB) during pulsed stress-method are evaluated using the following two procedures: 1) by adding up the total pulsed stress time, and 2) by extrapolation of the pulsed stress time to a constant voltage stress (CVS)-type measurements. It is shown that the latter method allows for a better comparison of identical oxides TDDB under various stress conditions. A methodology to characterize gate oxide breakdown using a single pulse is finally discussed. This is important to assess the gate-oxide failure condition during a charged device model (CDM)-type electrostatic discharge (ESD).

I. INTRODUCTION

Sub-nanosecond time-dependent characterization of gate oxide degradation and oxide breakdown mechanism are central to predicting and modeling oxide failure when a fast voltage transient is applied to a thin MOS gate-oxide. Understanding this phenomenon, on the other hand, is critical to developing guidelines and a systematic circuit design strategy that avoids common oxide damage resulting from Charged Device Model (CDM)-type electrostatic discharge (ESD) events [1]-[2].

CDM emulates the fastest ESD event, depicting a rise time and a total stress duration in the picoseconds and nanoseconds regime, respectively. This ESD event simulates the positive- or negative- charge built-up in the integrated circuit (IC) die and package through direct contact charging or through field induced charging [3]. The electrostatic discharge is generated via direct ground contact of one of the IC’s input/output (IO) pins. This ESD is very common in automated IC handling and assembly. Consequently it is nowadays widely considered to be the leading cause of ESD-induced IC failure.

The CDM waveform, however, is difficult to reproduce because it is highly dependent on the testing method, the die and the size of the IC package, among others [4]. To use a more controllable measurement strategy and be able to gain insight in the phenomena taking place in the CDM-time domain, Very-/Ultra- Fast Transmission Line Pulsing (VFTLP/UFTLP), with a pulse rise time in the hundreds of picoseconds and pulse width less than ten nanoseconds, is currently being used to assess CDM-like device response and CDM-induced failure. The standard TLP stress, on the other hand, provides a minimum of 200 ps rise time and 100 ns pulse width. To illustrate the difference between CDM and other ESD stress standards, Fig. 1 compares the most widely used ESD stress models waveforms as a function of time on a logarithmic scale. Notice that 1 ns (10^{-9} s) and 100 ns (10^{-7} s) correspond to approximately the pulse width of CDM and HBM, respectively, hence the use of standard TLP and VFTLP/UFTLP for current/voltage characterization in place of the HBM and CDM waveforms.

![Figure 1. Superimposed waveforms obtained for the standard ESD models (logarithmic time scale). Waveforms generated from SPICE simulation and IEC-1000-4-2 ESD standard [5]](image)

For a fast transient stress, such as CDM, the protection device may not respond fast enough to protect sensitive gate oxides and the time-dependent gate oxide breakdown voltage increases compared with the traditional DC voltage stress for long-term gate reliability evaluation. As a result, a different methodology is required to evaluate the gate reliability under CDM-type stress as well as ESD device transient response in the sub-nanosecond time domain.

The shortest time pulse width for CDM characterization reported in the literature is sub-50 ps [6] with a rise time in the picoseconds. Such pulses can also be generated during CDM stress by the inductive coupling between long metal lines.

Transmission line pulsed characterization of gate oxide has been widely published in the literature, [1, 7-14]. Standard TLP pulse is significantly wider in duration than a CDM-type ESD event and this measurement provides limited information on gate failure during a faster CDM event. Preliminary studies using VFTLP/UFTLP measurements to characterize gate...
degradation have been recently reported, [6], [10]. Notwithstanding, defining a predictive failure condition for circuit simulation or effective CDM ESD device protection remains elusive.

Two models have been proposed in the literature for the gate-oxide breakdown for thick oxide and under TLP stress conditions. The first is the power law model, [7], [12]-[13], [15]-[21], and the second is the 1/E model, [3], [8], [10], [22]. Recently the Hydrogen Release Model, [16]-[17], [23]-[26], linked to the power law model, has also been reported as the physical breakdown mechanism. Nonetheless it was refuted by several authors [27]-[30]. A major endorsement of the power law model, however, came from the experimental work performed by Nicollian et al. [31]. The latter showed that the time to breakdown did not reduce after decreasing the doping in the poly silicon gate.

The fit of experimental data to both the above models, and the connection of one possible physical explanation to the power law model gives the latter greater credibility. However, until the physical mechanism of oxide breakdown is well understood, the 1/E model cannot be discounted.

In this study, the experimental setup for the VFTLP measurements is reviewed. Next, the VFTLP waveforms close to the oxide breakdown condition are depicted and discussed. The power law model is then used to analyze the time dependent dielectric breakdown.

II. GATE OXIDE CHARACTERIZATION METHODOLOGY

A. Oxide Testing Procedures

A criterion to define the total stress time has been proposed by Wu et al. [10]. The total stress time was equated to the pulse width multiplied by the pulse count applied prior to the oxide damage, plus the time elapsed from the beginning of the pulse that damaged the oxide until the actual damage shown as an increase in leakage current. This methodology has the advantage of being straightforward to implement. However, as will be shown later in this study, this method does not allow for benchmarking of measurement results with other studies following different test procedures. Often the testing methodology leads to different interpretations and definitions of time to breakdown, further complicating comparison and interpretation of published experimental data.

There are five oxide testing methodologies considered in this study. The first is constant voltage stress (CVS), for which TDDB was initially defined, and is the time it takes from the instant the voltage is initially applied to an oxide until the time when the oxide is damaged (i.e. current flows through it). For this testing method, extensive experimental data is available.

A second closely related testing method is repetitive constant voltage stress (RCVS), in which TDDB is defined as the number of pulses applied, multiplied by the pulse width, plus the time to breakdown in the final pulse. It has been shown that this method reports longer time to breakdown. The time to breakdown under repetitive stress is a function of pulse width, frequency and stress voltage, showing the cumulative nature of the breakdown [7].

The third stressing methodology uses the previously discussed TLP- or VFTLP-type stress, commonly referred to as ramped voltage stress (RVS), where the voltage increases for each pulse [12]. In this method, it is difficult to define the time to breakdown as low voltage pulses do not have the same effect as the subsequent higher voltage pulses. As a result, if the time to breakdown is interpreted as the sum of all of the pulses [10], following the concept of repetitive CVS, it is dependent on the initial stress voltage and voltage step during stressing. In this study, a different method to transform the times using power law gives more importance to the later pulses and makes CVS comparisons possible. Both of these options are discussed in more detail later.

The fourth testing methodology is ramped voltage stress, which is done through constant current stress [32]. The influence of dV/dt can be studied in this way, and a model which can reproduce both CVS and CCS would likely be capable of predicting failure induced by arbitrary waveforms.

The final testing methodology consists of stressing the oxide with only one pulse of specific voltage magnitude, which needs to be found first experimentally, and record the time to breakdown within that pulse. It effectively provides a CVS-type result on a very short stress time. In this test method, however, a preliminary TLP-type measurement still needs to be performed to determine the actual pulse voltage across the oxide that can lead to failure. It requires an iterative process to identify this specific stress condition, but it can be further simplified once the specific oxide failure pattern is better understood.

The measurements in this study follow the third testing procedure. They are performed with two different waveforms, one at 1 ns pulse width and 100 ps rise time, corresponding to CDM-like stresses, and the other at 10 ns pulse width and 200 ps rise time, a pulse type that is about an order of magnitude wider than the previous one and in between the CDM-type VFTLP and the HBM-like TLP stress. The devices tested were thin oxide n- and p- MOSFETs in inversion mode. The gate oxide is formed via thermal oxidation, and the sample structures were tested at room temperature with the source and drain floating, which is a typical condition in the device during a CDM-type ESD event.

B. Experimental Setup and VFTLP Measurements Results

Figure 2 shows a high-level system diagram for the testing setup. It includes the charging of a transmission line by a high voltage source, which is discharged down the calibrated transmission line to stress the device under test, while voltage and current probes measure the device response. The measurements consist of recording the voltage versus time and current versus time waveforms for a variety of pulses, which are increased in energy for each pulse, and after each pulse a DC leakage current stress is performed.
The applied voltage step between the pulses is fixed, with the voltage stress increasing in steps of 0.5V per pulse starting at 0.5V until reaching failure. The leakage current measurement is performed after each pulse at 5V DC as depicted in Figure 3. To generate the classic TLP I-V characteristics, the voltage and current versus time pulse waveforms in the DUT are averaged between the 25% and 75% of the pulse durations to generate a single current-voltage (I-V) data point.

Oxide breakdown is observed at the subsequent pulse that damages the gate oxide. In the following discussion, it is detected by looking first at the point where the DC leakage current increases by orders of magnitude and later looking closer to the waveform where the hard failure occurred. Figure 4 shows three examples of fast-pulsed I-V characteristics where hard failure is detected at different levels. The oxide thickness affects the breakdown point in the I-V curve, making the breakdown voltage approximately 34.95V for the 70 Å thickness (Figure 4a) and 37.2 V for 130 Å (Figure 4b) for 1 ns duration pulses. In case of the longer pulses of 10 ns, the oxide breakdown occurs earlier (at 26.5V, see Figure 4c) due to the cumulative damage of longer pulses.

To illustrate the transient waveforms during different stress conditions, Fig. 5 shows the voltage and current versus time of the DUT at the breakdown point for two case study oxide thicknesses of 70 Å and 130 Å using a pulse of 100 ps rise time and 1.2 ns width. Additionally, the 70 Å gate oxide is also stressed using a 200 ps rise time and 10 ns pulse width signal. Figure 5a shows how the transient current waveform behaves increases abruptly in the Idut-Vdut characteristics in Fig. 4, where Idut-Vdut are calculated as the average of the current-voltage between 25% and 75% of the pulse duration. Once failure occurs and the voltage across the oxide drops as current flows through it, this average is no longer a valid measurement of the voltage that caused oxide failure. Notice that due to the statistical nature of trap generation, individual measurements may lead to seemingly contradictory results, as voltage distributions of reaching the critical trap density may overlap.
when the oxide hard-failure takes place. In this case, the current sharply increases to 0.14 A, approximately 2 ns after the initial part of the pulse. Figure 5b shows the current increase to 0.1 A, and in Fig. 5c the current increases first to 0.1 A, while voltages reach approximately 37.5 V, 39.5 V, and 32.2 V, respectively.

The breakdown point appears to be a transition point for the oxide from a reliable-to a damaged-insulator, as can be seen by how closely the breakdown waveform tracks the pre-breakdown and post-breakdown waveforms before and after this point, respectively. Notice in Fig. 6c that the breakdown current rises a second time far above the level it originally reached, but then returns to the lower average current level in the next pulse. This behavior has been observed before and its exact cause is not well understood, but different hypotheses have been discussed in reference [33].

III. STATISTICAL OXIDE STUDY

The statistical behavior of time-dependent dielectric breakdown can be characterized by the Weibull distribution [28]:

\[
F(T_{bd}) = 1 - e^{-\left(\frac{T_{bd}}{T_{63\%}}\right)^\beta},
\]

where F is the cumulative failure probability, the parameter \(T_{63\%}\) corresponds to the time when 63.2 % of the samples have failed, and the shape parameter \(\beta\) represents the slope in a graphical representation of the Weibull distribution:

\[
\ln[-\ln(1 - F(T_{bd}))] = \beta \ln\left(\frac{T_{bd}}{T_{63\%}}\right)
\]
The same equations can be applied likewise using the charge to breakdown, $Q_{BD}$, replacing the time to breakdown as the independent variable [15]. $Q_{BD}$ is thought to be more closely related to the physical mechanism of breakdown [15].

Area dependence can be obtained by using Equation 3 [7], [17], [30]:

$$T_{BD}(A_2) = T_{BD}(A_1) \left( \frac{A_1}{A_2} \right)^{\frac{1}{\beta}}$$

(3)

Area dependence is observed as a vertical shift in the Weibull plot [15]. Assuming that the failure probability of a specific oxide area is $p$, then the failure function of the multiple $N$ independent areas failing is $F=1-(1-p)^N$.

**IV. POWER LAW MODEL**

In constant voltage stress (CVS) tests, empirical TDDB models have been proposed to easily estimate the breakdown performance of a characterized oxide. Two of these models are well known, the exponential law and the power law [15], [17].

The TDDB power law is defined in (4) as, [15] and [3]:

$$t_{bd} = a V_G^{-n}.$$  

(4)

The TDDB exponential law is defined as [17]:

$$t_{bd} = t_0 e^{-\gamma V_G},$$  

(5)

where $\gamma$ is the voltage acceleration factor.

It was recently demonstrated that the power law more accurately predicts TDDB than the exponential law [15], [17], [20]. The power law’s exponent, $n$, is independent of oxide thickness and the value for both n- an p- MOSFET is ‘44’ [15], hence it is more predictive.

**V. STATISTICAL GATE OXIDE EVALUATION UNDER VFTLP AND DISCUSSION**

Both the Maximum Likelihood Method (MLE) and Rank Regression in X methods were used to obtain the values of $\beta$ in this study. Fig. 7 depicts the Weibull plots for n- and p- MOSFETs using the simple time to breakdown definition, and a stress condition of 100 ps rise time and 1 ns pulse duration for oxide area ($A_{ox}$)=30 $\mu$m$^2$ and oxide thickness ($t_{ox}$)=70 Å.

Figure 8 shows the Weibull plot for the case of 200 ps rise time and 10 ns pulse duration. The time to breakdown increases and the voltage to breakdown decreases due to the longer pulse times, while the $\beta$ differ slightly from those in Fig. 7 for n- and p-MOSFETs, respectively. For a thicker oxide $t_{ox}$=130 Å and same area, it is observed that the oxide lifetime and $\beta$ values increase compared to the thinner oxide [34]-[35], as depicted in Fig. 9.

Figure 10 shows the Weibull plot for NMOS devices with different areas, where one is twice the other. Notice the area dependence in terms of the $T_{63\%}$ in which the larger area device has a lower $T_{63\%}$ time. To have a more comprehensive evaluation of the area-dependence, a larger set of -samples and -area difference need to be evaluated.

The Weibull slopes extracted using the simple TDDB definition provides the $\beta$’s that do not correlate with the values obtained during CVS measurements [1], [7], [12], [15], [28], [36]-[40]. However, the plots previously discussed depict similar tendencies.
The power law definition and its constant $n = 44$, [15], lead one to the conclusion that the VFTLP pulses prior to the final pulse that destroys the oxide are less significant. Therefore, an alternative mathematical procedure to interpret the TDDB results under very fast pulse testing consists on: 1) treating each VFTLP pulse as a CVS stress of a very short time, 2) transform the time to breakdown to a fixed voltage using an alternate form of the power law, shown in Equation 6 below, and 3) sum all of these times together for each device.

$$t_{bd2} = t_{bd1} \left( \frac{V_f}{V_i} \right)^n$$  \hspace{1cm} (6)

For the train of very fast transmission line pulses, the total pulse time and the average pulse voltage can be used in Equation 6 for $t_{bd1}$ and $V_i$, respectively. However, for the final pulse in which the oxide actually fails, $t_{bd1}$ and $V_i$ correspond to the part of the pulse prior to oxide breakdown. Specifically, $t_{bd1}$ is defined as the difference between the beginning of the stress period, when the pulse’s initial rise-time is completed, and the failure time, when the current begins its rise above its previous average within the pulse. On the other hand, $V_i$ is determined by starting from the voltage at the failure time found above, moving backwards in time ten data points and then averaging the eleven voltage points surrounding this point (the point, five behind, and five ahead), except when the failure time is too close to the beginning of the stress period, in which case the points between the beginning of the stress period and the failure time are averaged. In order to sum the times, they must all be on the same voltage scale, thus making $V_i$ a constant.

Figure 11 shows a CVS-type $\beta$ value Weibull plot when the time to breakdown is extrapolated using a fixed voltage of 20V. The value of 20V was chosen arbitrarily to some extent, as the distribution showed no differences in $\beta$ for the chosen voltage; only values of TDDB shifted up or down exponentially. The choice of 20V, which is a voltage below the failure voltages measured during the different stress conditions, places the calculated TDDBs in a time range above that of CDM, which allow for direct comparison with stresses of equivalent stress times.
tox=70 Å Aox=40.5 μm² tr=100 ps tp=1 ns

Figure 12. Weibull plots for 40.5 μm² NMOS for 100 ps rise time and 1 ns pulse duration for tox=70 Å, with (A) 0.5 V starting pulses and (B) Mixed 0.5 V and 15 V starting pulses, both using the simple TDDB definition.

The power law transformation of the time to breakdown introduces a nearly undetectable change in β, once the different testing step results are added and the time to breakdown scale. Notice, however, the transformed data in Fig. 13a and 13b did not fit to the linear extrapolation as well as the untransformed data in 12a. There are two reasons for this: First, the early pulses that cause low-stress to the oxide are given the same weight as the latter high-stress pulses in the simple TDDB definition, so they act as a padding that increases the β value by reducing the percentage difference between failure times making the fit in 12a unrealistically high. Second, the measuring equipment has an error margin for determining the time the current or voltage is at a certain level and this Δt is also amplified by the power law, making the time to breakdown values less certain for the higher voltage pulses, increasing the TDDB scatter.

This transformation concept to convert standard TLP-type stressing into CVS, therefore, is not proposed to replace real CVS measurements for determining long-term oxide breakdown characteristics. Instead, it can make it possible to compare breakdown data across many stress regimes and conditions. As an example, Fig. 14 shows the Weibull plot when 5 types of stressing were applied to the same 40.5 μm² NMOS, including the stress types covered earlier with a pulse width and rise time of 100 ps, 1 ns and 200 ps, 10 ns respectively, and the case where the starting pulse voltage for some 100ps, 1ns NMOS is set to 15V, as well as two new cases of 200ps, 1ns and 400ps, 1ns.

Figure 13. Weibull plots for 40.5 μm² NMOS for 100 ps rise time and 1 ns pulse duration for tox=70 Å, with (A) 0.5 V starting pulses and (B) Mixed 0.5 V and 15 V starting pulses, both using the power law transformation.

Figure 14. Weibull plot for tox=70 Å Aox=40.5 μm² NMOS for five stress regimes in rise time (RT), pulse width (PW), and initial stress voltage (Vo) where the VFTLP measurement starts.

Under these different measurement conditions, it still produces roughly the same β and time domain from the fitted curve regardless of the stress differences. Notice that the TDDB points for each measurement methodology are mixed together, rather than remaining in distinct groupings, indicating
that they are indistinguishable after transformation and that the comparison is a valid one.

VI. CONCLUSIONS

VFTLP (RVS) measurements have been performed on gate oxide and the statistical behavior of the time to breakdown has been obtained in two different ways: first by adding up the total time of stress, and second by transforming it into CVS time to breakdown. Further study needs to be performed to more accurately obtain the transformation to CVS stress and correlate the reported CVS β parameter with the VFTLP β parameter, as the transformation yielded slightly lower than expected β’s, and the untransformed data’s impossibly high β’s. Additional study in which the oxide is stressed with one pulse only in which the leakage current change indicative of breakdown occurs during that pulse needs to be performed. Once the three types of measurements, CVS, RVS, and one-pulse CVS, are correlated the evaluation of time to breakdown is expected to be defined more systematically and oxide characterization time considerably reduced.

ACKNOWLEDGEMENTS

Thanks to Alan Righter, Dave Clarke, Srivatsan Parthasarathy, Bruce Atwood, Thorsten Weyl, and Andrew Olney for valuable discussion and support to this project.

REFERENCES