Soft breakdown triggers for large area capacitors under constant voltage stress

J. Schmitz, H.J. Kretschmann, H.P. Tuinhout, and P.H. Woerlee
Philips Research Laboratories
Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands
Tel. +31-40-2742723; Fax: +31-40-2744113; E-mail: Jurriaan.Schmitz@philips.com

ABSTRACT

This work quantitatively compares breakdown triggers for constant voltage stress of large area NMOS capacitors (up to 10 mm^2) with 1.8 to 12 nm gate oxide thickness (with negative V_G). We conclude that in the studied range, breakdown is identified more reliably with a current step trigger than through increased current fluctuation (RMS). We also present data filtering algorithms that significantly enhance the ratio between the breakdown signal and background noise level.

INTRODUCTION

Monitoring the extrinsic defect density of the gate oxide using constant voltage stress (CVS) and constant current stress (CCS) has proven to be a useful technique to maintain high yield and reliability for the manufacturing of CMOS IC's. These types of stress tests are generally part of a more complete reliability investigation, such as described e.g. in [1]. For extrinsic defect density evaluation in a mature process, it is a prerequisite that a relatively large total capacitor area is tested, of the order of a few square centimetres per wafer. This is traditionally done using large area capacitors (~1-10 mm^2).

The continuing trend of downsizing the gate oxide thickness however severely complicates such measurements, because electrical breakdown becomes soft [2-5] while the tunnel current before breakdown increases exponentially for thinner layers. As a result, it becomes difficult to identify a dielectric breakdown on a large area capacitor when the dielectric thickness is less than 5 nm.

In this paper, we quantify trends of dielectric breakdown signatures as a function of oxide thickness and area. We show that extrinsic defect density determination can be approached conventionally (by adding instrumentation knowledge and enhancing the data analysis techniques) down to around 2.5 nm gate oxide thickness; whereas beyond that point, a new approach is required.

This paper discusses the signature of the first breakdown occurring on a device under test, irrespective of its mode (soft or hard). Soft and hard breakdowns are treated equally in the analysis.

TRIGGER REQUIREMENTS

A breakdown trigger is supposed to detect a breakdown of a monitored capacitor. For the specific purpose of a CVS or CCS test, it should at least fulfil the following requirements:

- It should return the breakdown time t_{BD}, with a sufficiently high resolution (the relative error on the time should be smaller for steeper Weibull distributions).
- It should 'never' generate false triggers. A false trigger may be misinterpreted as an early breakdown, thereby suggesting an extrinsic defect. This would turn a measurement problem into a supposed fabrication problem. For most purposes, the false trigger rate must be below 1 in 100 measurements.
- It should 'never' miss a breakdown event. The demands to this respect are however less stringent than on the false trigger rate.
- The false trigger probability and the missed breakdown probability should be constant over time, so that the risk of an erroneous trigger is spread equally over the time of the measurement. This requirement is very hard to meet.
- The trigger should be real-time and fast.
- It should be easy to understand, for interpretation of deviating results.
- Its implementation should be independent of device properties (such as oxide thickness and capacitor area) and the test platform.

Additional requirements may be present for specific circumstances. For instance, it may be interesting to store the breakdown signature and the SILC after breakdown, and possibly to continue the test until hard breakdown is observed. Throughout this work, we constrained ourselves to triggers generated in software, intended to run real-time during the measurement.

EXPERIMENTAL

The measurements are performed on NMOS capacitors with drain diffusion edges, fabricated in a stripped 0.18 µm dual-flavoured CMOS process with titanium silicide and single Al metallization in a 6" pilot line. The (pure SiO_2) gate oxide thickness was varied between 1.8, 2.5, 3.3, 4.0, 5.5, 7.0, 9.0 and 12.0 nm (ellipsometric values), while all other process variables were kept constant. The capacitors were tested with $V_D = V_B = 0$ V, the gate voltage forced, while the gate current was measured repetitively until a predefined time was reached. All raw current measurements were saved to disk. Breakdown evaluation was done off-line, enabling the comparison of different breakdown triggers on identical data sets. We used a Keithley S400 parametric tester with an Electroglas probe station.

RESULTS

1. Breakdown signal decrease

In essence, breakdown detection boils down to the determination of the point in time where the conducting properties of the device under test change. Many approaches can be used for this goal. We have limited ourselves to a very simple one: to repetitively measure the current during a constant voltage stress. Other approaches that can be followed include the use of a dedicated noise characterization setup, or by periodically interrupting the stress bias to measure the low-voltage SILC. We followed the arrangement of the conventional test, to find out what its final limitations are.

The magnitude of breakdown signals, defined as the relative current increase ΔI/I during CVS or the relative voltage drop ΔV/V during CCS, decreases monotonously as a function of oxide thickness, as demonstrated in Figure 1. Traditional breakdown criteria therefore rap-

idly lose significance. For instance, a 20-50% voltage drop as required in many conventional CCS tests only works for oxides thicker than 4-6 nm (depending on area), and the commonly applied 10 mA/cm^2-criterion [1] in CVS tests is obviously wrong for thin oxides where the tunnelling current may exceed this threshold from the start. Hence, a more sensitive, and more precise breakdown criterion must be defined.

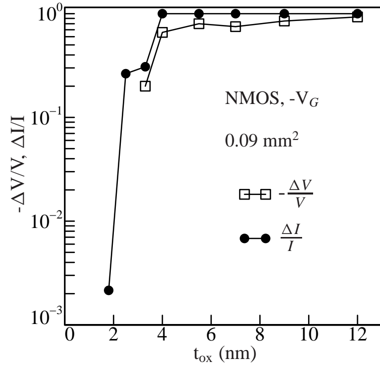

Figure 1: Typical (i.e. median) voltage drop and current increase in constant current and constant voltage stress, respectively, as a function of oxide thickness. The current increase was limited to 100% by the current compliance. The values are medians of 30 tested capacitors.

With decreasing breakdown signals, the background becomes important: when triggering on weak signals, spurious effects could be misinterpreted as a breakdown. Examples are system noise, device noise, and other disturbances of the measurement (e.g. from electromagnetic interference). Figure 2a shows an example of the current development with time of a capacitor under CVS. A breakdown trigger may for instance be monitoring the RMS variation of the current (Figure 2b), or the current increase ΔI (figure 2c). Evidently the trigger observables ΔI/I and RMS/I are not zero before breakdown, predominantly because the measurement contains (white) noise. However, in this particular example the breakdown manifests itself clearly in both the RMS/I and the ΔI/I graph by a sudden sharp increase. Automatic triggering on this particular breakdown event should not be difficult, but a threshold level needs to be defined that separates the breakdown signal from background noise.

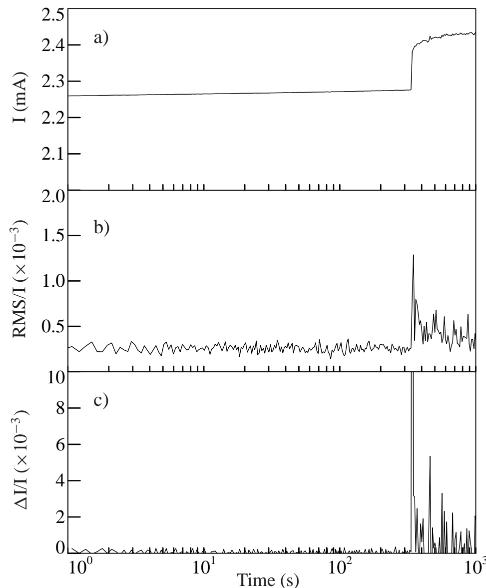

Figure 2: Current as a function of time, and the derived (relative) RMS variation and ΔI/I, on a 2.5 nm t_{ox}, 0.81 mm^2 area capacitor under constant voltage stress of -4.0 V. The soft breakdown (at t = 341 s) is clearly seen in all three curves.

2. Data handling

In this work, we have considered a large number of I-t data sets. We first composed a reference table for the times-to-*first*-breakdown (t_{BD}) of each capacitor. This was done using a simple, oversensitive ΔI/I trigger; fake triggers were manually corrected using inspection of the I-t curves and derivates by eye. Once each time-to-breakdown is determined, two time windows are defined: the "pre-breakdown" time window, which ends at $0.95 \times t_{BD}$, and the "breakdown" window, that runs between $0.95 \times t_{BD}$ and $1.05 \times t_{BD}$ (see Figure 3). In the pre-breakdown window, we determine the highest peak of a given trigger observable. This is the worst-case jump of the background, an important entity that tells us how bad the background can be. In the breakdown window, we also determine the highest peak of the given trigger observable: this is assumed to be the identifier of the breakdown.

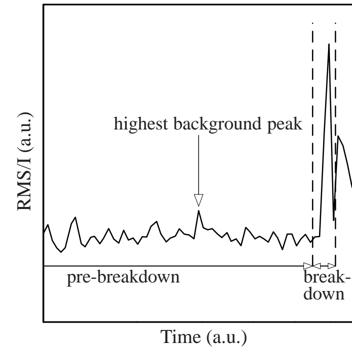

Figure 3: Development of an observable (RMS/I) over time; and the pre-breakdown and breakdown windows, in which the highest value of RMS/I is searched as indicators of "background" and "breakdown signal" respectively.

Doing this exercise on 30 capacitors with equal design and equal bias conditions results in a set of 30 "background peak" values and 30 "breakdown signal" values. An example of the distribution of these values is shown in Figure 4. (The Weibull representation of these data is chosen for convenience; it is not meant to suggest Weibull statistics.) From each of these sets, we determine the median ("typical"), as well as the worst-case values. These values are used as characteristic values for the distribution throughout this work.

The approach sketched above enables quantitative comparisons of various trigger approaches. The distributions of breakdown signals and background noise peaks should be well separated to be able to trigger on each breakdown, without generating fake triggers on the background. In other words, the weakest breakdown signal should be well above the highest background peak. A trigger threshold can be chosen freely between the highest background peak and the lowest breakdown signal.

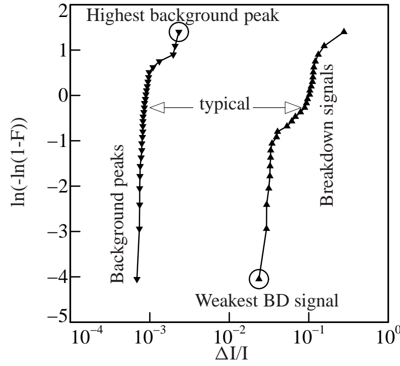

Figure 4: Distribution of breakdown signals and background peaks for a ΔI/I trigger (29 capacitors with t_{ox}=2.5 nm, 0.81 mm^2, V_G = -4.0V, N=16). Indicated are the key entities used throughout this work: the *typical* background peak and *typical* breakdown signal (being the median values from these distributions), as well as the highest background peak and the weakest breakdown signal. One of the 30 tested capacitors showed early failure and was skipped.

3. Data filtering algorithms

Because the breakdown signal and the background tend to come close together, we incorporated data filtering algorithms to improve the signal-to-noise ratio. Careful analysis of the collected data has shown that false triggers are usually set off by single outlying data points, causing an incidentally large ΔI/I and RMS/I value. We therefore follow a simple smoothing procedure:

- Consider N consecutive data points (N typically 16);
- Compute the average value of these N data points;
- Optionally: remove M data points that are farthest away from their average (M typically 2);
- Compute the average value of the remaining N-M data points.

The resulting average value is symbolized I(t), where t is the time at which the first of N data points was measured. Similarly, the root-mean-square variation of the same group of N-M data points is symbolized by RMS(t). We analysed triggers of the type

$$\frac{\Delta I}{I} = \frac{I(t+dt) - I(t)}{I(t)} > T_1$$

$$\frac{\text{RMS}}{I} = \frac{\text{RMS}(t)}{I(t)} > T_2$$

where T_1 and T_2 are fixed threshold levels, and I(t+dt) is computed from the N data points immediately following the data points used for the computation of I(t). In this work, we apply the same data filtering on I(t+dt) as on I(t).

Taking averages over N data points should result in a decrease of the background with $1/\sqrt{N}$ when the measurement noise is truly white. The observed dependence (Figure 5) is weaker than that, but still significant. As can be expected for a more or less step-like change of the conductivity of the capacitor (like in Figure 2), the breakdown signal is insensitive to averaging.

Removing 1, 2 or 3 extreme values from each set of N=16 data points hardly affects the signal and background (Figure 6), and this method can therefore be used without penalty to become less sensitive to single outlying data points. The positive effect of extreme value removal on a data set with disturbances is exemplified in Figure 7. We encountered several events like this in our study (be it with an occurrence rate of less than 1% per stress test), but have not identified the cause of the disturbance that leads to the transient outlying measurement values. We expect that random outlyers like the one shown in Figure 7 are even more abundant in a (usually noisy) fab test area.

When averaging and extreme removal are both applied, a ΔI/I trigger correctly identifies all breakdowns at 2.5 nm gate oxide when the trigger levels are appropriately set. For the largest area capacitors (10 mm^2), the selection of the trigger threshold level is critical: see Figure 8. In this figure it is clearly seen that the *relative* current jump is proportional to the inverse of area, which supports the model that the breakdown spot conductivity is independent of the capacitor area.

From this study we concluded that triggering on a soft breakdown is easily achieved with the ΔI/I trigger that includes data point grouping, averaging and removal of extremes. Only the largest area capacitors (10 mm^2) have breakdown signals very close to the background.

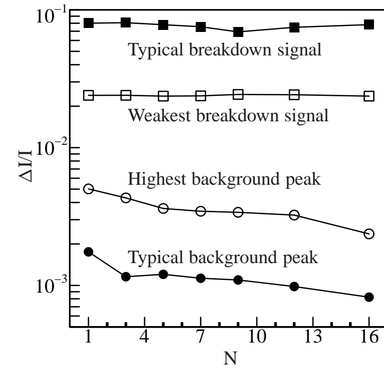

Figure 5: Current increase signal and background, as a function of the group size N of clustered data points. A larger group size results in background suppression (with a factor 2.1 between N=1 and N=16) and therefore the (unaffected) breakdown signal is more pronounced. Data from 29 capacitors with V_G = -4.0 V, t_{ox} = 2.5 nm, A = 0.81 mm^2.

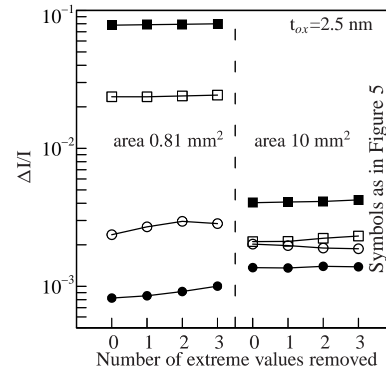

Figure 6: Breakdown signal and background peaks of 16-point averages, and after exclusion of 1, 2, and 3 extreme values. Symbols are as in Figure 5. Removal of extremes results in a slight (insignificant) reduction of the margin between signal and background on the 0.81 mm^2 capacitors, and on a slight improvement on the 10 mm^2 capacitors (both with a factor 1.2).

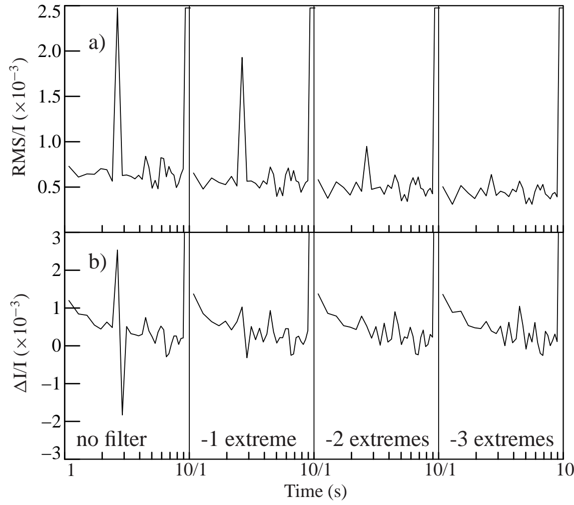

Figure 7: Development of a) RMS/I and b) ΔI/I with time in case of outlying data points. At t=3 seconds, a disturbance occurs; at t=9s, breakdown is observed. The figure shows the 16-point values of RMS/I and ΔI/I with time, as well as the 15, 14, 13 point values after exclusion of extreme values from each set of 16 points. The disturbance spike is filtered out by extreme value removal.

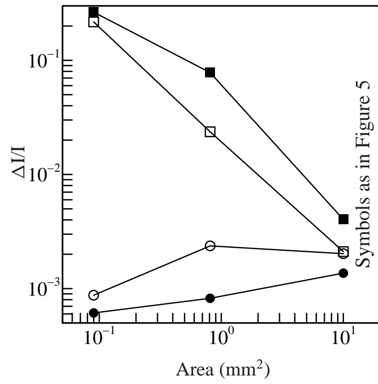

Figure 8: Current increase after breakdown as a function of the capacitor area, with t_{ox} = 2.5 nm, V_G=-4.0 V. Symbols as in Figure 5; 16-point averages are used for ΔI/I computation. ΔI/I decreases linearly with area, while the background increases slightly.

4. Noise triggers

Increased noise in the gate current is commonly observed after soft breakdown [2,7] and was proposed to be used as an identifier for breakdown [6,8]. Instead of considering dedicated noise measuring hardware, we estimated the noise from the available set of DC measurements, by comparing consecutive data points like proposed by G. Alers et al. [6] for J-ramp stress. Note that RMS convergence is slower with N than the convergence of the mean value I, and therefore one requires a larger number of data points if an accurate estimate is needed.

The RMS trigger signal and background are shown in Figure 9. Signal and background are dramatically less separated than the same parameters for a ΔI/I trigger on the same set of data (cf. Figure 6). The weaker breakdown signature of RMS/I has been confirmed on all other areas, as well as on thinner and thicker oxides. Moreover, we have also tested under constant current stress, and under comparable circumstances the voltage drop ΔV/V gives a more pronounced signal than the RMS variation on the voltage. This important result is the first published quantitative comparison between a noise trigger as described in [6] and a conventional trigger, and it contradicts an earlier (qualitative) observation that the only reliable way to detect soft breakdown of ultra-thin oxides is to monitor the noise on I_G or V_G [8].

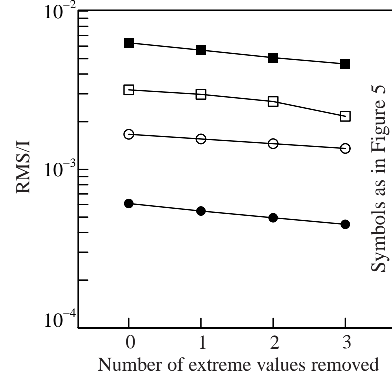

Figure 9: Breakdown signal and background of the RMS variation on the current for capacitors with t_{ox} = 2.5 nm, V_G=-4.0 V, A=0.81 mm^2 (as in Figure 6). The figure shows the 16-point RMS, and the values after exclusion of 1, 2 and 3 extreme values.

We expect that the superiority of the ΔI/I signal is not strongly platform dependent. A more accurate and better shielded measurement system may suppress the background levels, but this will to first order hold equally for the ΔI/I background and the RMS background, since they are both dominated by short-term repeatability errors [9].

The difference between signal and background presented in Figure 9 is so small (less than a factor 2), that we cannot tolerate a decrease of N below 16 in this case. This is demonstrated in Figure 10, where N is varied from 5 to 16. The margin between signal and background improves only slowly, as expected. Beyond N=16, the signal to background ratio will still improve.

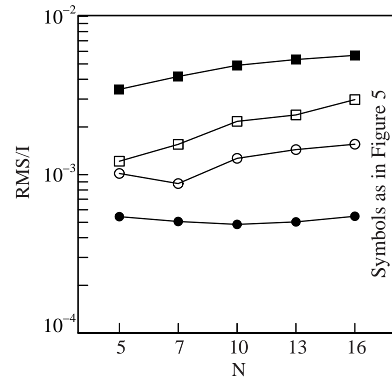

Figure 10: Breakdown signal and background of the RMS variation on the current for capacitors with t_{ox} = 2.5 nm, V_G=-4.0 V, A=0.81 mm^2, as a function of the group size N of data points used for the RMS computation. One extreme value is removed from each group. The difference between signal and background gradually increases with larger group size N.

5. A combined noise/current jump trigger

Where triggering on a current jump and a noise increase becomes difficult due to poor signal-to-background ratio, it might be interesting to use a combined trigger, i.e. one that triggers on a current jump, but will also trigger on a noise increase when no significant current jump is present. It was found, however, that such an approach does not improve the trigger performance, because weak RMS/I signals coincide with weak ΔI/I signals at breakdown. This is shown for a variety of capacitor areas and for 1.8 and 2.5 nm oxide thickness in Figure 11. Note that the typical background level for RMS/I is around 6×10^{-4} and for ΔI/I around 10^{-3}. These levels are however not indicated in the figure because they vary slightly for different stress currents (as seen in Figure 8, and discussed in detail in [9]). As should be expected from the strong correlation depicted in Figure 11, indeed we observe that a combined ΔI&RMS trigger does not give an improved signal-to-background ratio. Figure 11 illustrates again that the ΔI/I signals are typically a factor 3-10 stronger than the RMS/I signals.

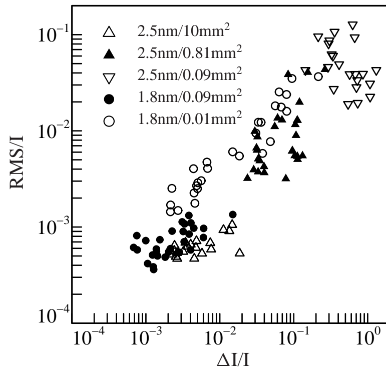

Figure 11: Correlation between RMS/I signal (N=16) and ΔI/I signal (N=16) of breakdowns on capacitors with t_{ox} = 1.8 nm (circles; stressed at -3.9 V) and 2.5 nm (triangles; stressed at -4.0 V) and various areas. Weak ΔI/I signals coincide with a small increase in the RMS/I variation. Those capacitors with post-breakdown RMS/I close to background still show significant ΔI/I.

6. Triggering at 1.8 nm

With thinner gate oxide, the 'fresh' tunnel current exponentially increases, blurring the signature of a breakdown spot. As a result, we rapidly lose the breakdown signal on large area capacitors below 2.5 nm gate oxide thickness. Figure 12 shows the performance of the ΔI/I trigger and some alternative triggers, on a 0.09 mm^2 area capacitor with t_{ox} = 1.8 nm gate oxide. (Note that we analysed data from smaller area capacitors, because on 0.81 mm^2 and larger areas, we were not able to identify a breakdown signal.)

The figure shows that for the previously described triggers, the highest background peak is higher than the weakest breakdown signal; therefore in their current implementation, they fail to work in a foolproof manner (they will either generate false triggers, or miss real breakdowns).

With the ΔI/I trigger, including more data points into the analysis gives some remorse, as illustrated by the trigger results in Figure 12. The triggers of type '*n*+*m*' compare the highest I-value from *n* consecutive groups before breakdown, to the lowest of *m* groups after breakdown. In this notation, '1+1' is the previously discussed ΔI/I trigger. '2+2' can be written as follows:

$$I_{\text{before}} = \text{MAX}(I(t-2dt), I(t-dt))$$

$$I_{\text{after}} = \text{MIN}(I(t), I(t+dt))$$

$$\frac{I_{\text{after}} - I_{\text{before}}}{I_{\text{before}}} > T_1$$

using the same notation as in subsection 3. (If this condition is met, breakdown occurs at time t.) With *n*>1, isolated low values of I cannot cause a trigger; with *m*>1, the trigger becomes insensitive to isolated high values of I. Going from the standard '1+1' trigger to the '1+2' trigger, and hence including 16 more data points to the trigger analysis, we obtain a better separation between signal and background: see Figure 12. Further increase of *n* or *m* has no effect at all, as is clear from the figure. The '1+2' trigger on these capacitors is the best trigger we evaluated in this work.

However, note that this result relates to a relatively small area capacitor: the signal-to-background drops dramatically from 2.5 to 1.8 nm, and also with increasing area, while this type of data filtering only incrementally improves the signal-to-background ratio. We can therefore conclude that for large area capacitors having t_{ox}=1.8 nm or less, this TDDB measuring approach is no longer valid.

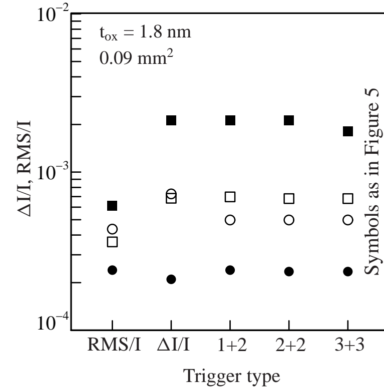

Figure 12: Results of the standard RMS/I and ΔI/I triggers on 30 capacitors with t_{ox} = 1.8 nm, A=0.09 mm^2, V_G=-3.9V; using N=16. These two triggers have a significant failure rate: the worst-case background (open circles) exceeds the weakest breakdown signal (open squares). Also shown are the signal and background for ΔI/I-like triggers involving more than two groups of 16 data points, symbolized with '1+2', '2+2', and '3+3' (see text).

7. Impact of measurement integration time

In this study we applied the shortest integration time possible on our instrument, resulting in the best achievable time resolution. A longer integration time is expected to give more accurate data points, and therefore may improve the margin between breakdown signal and background. We ran a few tests with a longer integration time (1 power line cycle (PLC), 20 ms, as opposed to 1 ms for the other tests). Indeed we observe some improvement of the noise floor, as seen in Figure 13. However, we pay a high penalty: first of all, the background increases strongly in the first seconds, because the sampling time scale comes closer to the time scale of charge trapping. This can be clearly observed in the figure. As a result, the trigger sensitivity is poor during the first seconds, which is unacceptable in view of the trigger requirements mentioned before. Secondly, we have observed that the occurrence of

incidental spikes (single outlying data points) increases strongly when the 1 PLC integration time is used. We conclude that the shortest integration time is the most practical.

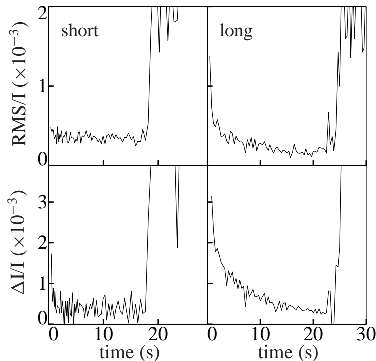

Figure 13: Effect of the integration time on the evolution of ΔI/I and RMS with time, qualitatively illustrated on typical examples. A ‘long’ (20 ms) measurement integration time as opposed to standard (’short’, 1 ms) gives rise to a high background in the first seconds, due to charge trapping in the capacitor. N = 16, V_G = -4.0 V, t_{ox} = 2.5 nm and A = 0.81 mm^2.

8. Gate voltage dependence

All results reported in the previous sections were obtained at a fixed gate voltage. It is important to verify the validity of the drawn conclusions under different bias conditions. From stress tests at various gate voltages, we compiled Figure 14. It shows the ΔI/I at breakdown, and the RMS/I at breakdown, for 6×15 capacitors in Weibull representation. It shows that the two breakdown signatures, relative noise increase and relative current increase, are hardly dependent on the applied stress bias. Only at the most severe bias condition (-4.1 V) a significant decrease in the RMS/I and ΔI/I breakdown signals is observed. The system noise background is to first order independent of gate voltage when it is normalized to I in the same way, implying that the signal-to-background ratio shows no significant bias-dependence in the studied range of V_G. This important result shows that the observed trends presented in this work are not restricted to a particular bias condition but rather describe the situation for constant voltage stresses in the entire range normally used for constant voltage testing.

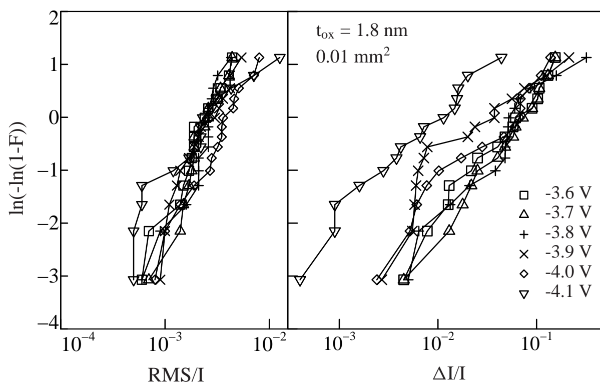

Figure 14: Distribution of RMS/I (left) and ΔI/I (right) values after breakdown, for several gate bias conditions, on 0.01 mm^2 capacitors with t_{ox} = 1.8 nm (N=16). The gate bias has no significant influence on the breakdown signal distribution, except in the highest V_G case (-4.1 V), where the ΔI/I and RMS/I signals are rapidly decreasing.

CONCLUSIONS

We present a thorough systematic, quantitative comparison of schemes for breakdown detection in a CVS test. A current step trigger (on ΔI/I) is well applicable on large area capacitors down to t_{ox}=2.5 nm when data filtering is applied. Noise triggers (on RMS) are significantly less sensitive to soft breakdown, contrary to earlier reports. Triggering on RMS noise definitely does not solve the problem of a weaker ΔI/I breakdown signature on large area capacitors. On 1.8 nm thick oxides conventional triggering becomes impossible for > 0.1 mm^2 capacitors, due to the large tunneling current. More advanced trigger algorithms on the same data can only yield incremental improvements. Therefore, a new approach is required for extrinsic defect monitoring on sub-2 nm oxides. These conclusions are shown to be independent of the gate bias, while the measurement integration time slightly affects the system background, favouring short integration times. This work compiles a large amount of quantitative results obtained on large area capacitors under constant voltage stress, and our conclusions are of vital importance for extrinsic defect density determination in present and future CMOS generations.

ACKNOWLEDGEMENTS

We would like to thank FAB-WAG, Anja van Brandenburg and Josine Loo for sample manufacturing, Adrie Zegers-van Duijnhoven for her spectral analysis work, and Hein Valk, Fred Kuper, Johan Klootwijk, Robin Degraeve, Benjamin Kaczer and Philippe Roussel for fruitful discussions. This work was supported by the EC project HUNT.

REFERENCES

[1] G. Ghibaudo et al., *Accelerated dielectric breakdown and wearout standard testing methods and structures for reliability evaluation of thin oxides*, Microelectronics Reliability 39 (1999) page 597.

[2] S.H. Lee et al., *Quasi-breakdown of ultrathin gate oxides under high field stress*, IEDM Technical Digest page 605, 1994.

[3] E. Wu et al., *Structural dependence of dielectric breakdown in ultrathin gate oxides and its relationship to soft breakdown modes and device failure*, IEDM Technical Digest page 187, 1998.

[4] M. Alam et al., *Explanation of soft and hard breakdown and its consequences for area scaling*, IEDM Technical Digest page 449, 1999.

[5] M. Depas et al., *Soft breakdown of ultra-thin gate oxide layers*, Transactions on Electron Devices vol. 43 no. 9 page 1499, 1996.

[6] G.B. Alers et al., *J-ramp on sub-3 nm dielectrics: noise as a breakdown criterion*, IRPS proceedings, Vol. 37 page 410, 1999.

[7] G.B. Alers et al., *Tunneling current noise in thin gate oxides*, Applied Physics Letters 69 (19) page 2885, 1996.

[8] B. Weir et al., *Ultra-thin gate dielectrics: they break down, but do they fail?*, IEDM Technical Digest page 73, 1997.

[9] J. Schmitz and H.P. Tuinhout, *A study of measurement system noise for sensitive soft breakdown triggering*, ICMTS proceedings, Kobe 2001 (in print).