An Array-Based Test Circuit for Fully Automated Gate Dielectric Breakdown Characterization

John Keane Shrinivas Venkatraman Paulo Butzen* Chris H. Kim

*State University of Rio Grande do Sul, Porto Alegre, Brazil
University of Minnesota, Minneapolis

***Abstract*- We propose an array-based test circuit for efficiently characterizing gate dielectric breakdown. Such a design is highly beneficial when studying this statistical process, where up to thousands of samples are needed to create an accurate time to breakdown distribution. The proposed circuit also facilitates investigations of any spatial correlation of dielectric failures, and can monitor a progressive decrease in gate resistance. Measurement results are presented from a 32x32 test array implemented in a 130nm process.**

I. Introduction

While scaling CMOS device dimensions allows designers to pack more, and faster, transistors on a die, it also leads to an increased susceptibility to variations and reliability mechanisms. One such reliability issue is time-dependent dielectric breakdown (TDDB). This mechanism causes a conduction path to form through a gate dielectric layer placed under electrical stress, leading to parametric or functional failure. Breakdown has been a cause for increasing concern as gate dielectric thicknesses are scaled down to the one nanometer range, because a smaller critical density of traps is needed to form a conducting path through these thin layers, and stronger electric fields are formed across gate insulators when voltages are not scaled as aggressively as device dimensions.

Although many of the physical details behind TDDB are still under debate, the percolation model is widely used to describe the gradual accumulation of electrical defects through a stressed oxide, which eventually form a current conduction path resulting in the breakdown [1]. Some studies have used the time to the first breakdown event (defined as an increase in gate current to some pre-determined level) to extrapolate predicted device lifetimes from accelerated stress experiments [2]. A range of currents can be detected after this first event, and the distinction between current paths with low and high conduction levels led to the classification of "soft" and "hard" breakdowns. However, the definitions of those terms are contentious, and some authors claim that all breakdown events are more correctly described as progressive in nature [3]. In addition, it has become apparent that transistors can continue to function in certain cases after that first breakdown event, and the progressive, post-breakdown current evolution must also be taken into consideration to obtain a less pessimistic lifetime projection [3-5]. This is particularly true when operating at lower voltages and with thinner dielectrics, making an observable progressive breakdown current more likely before final device failure [5].

TDDB is a function of a number of variables, including the gate voltage and oxide thickness as mentioned earlier, as well as temperature, device area, and dielectric materials and purity. Several models have been used to describe the relationship between the time to failure due to breakdown and these variables, but additional work is needed to more fully characterize TDDB in general so that the correct predictive models can be selected. The breakdown behavior of each new CMOS process must also be thoroughly tested during the process characterization phase in order to obtain a detailed understanding of the technology reliability.

In this paper, we present a circuit design that performs automated measurements in a test array to efficiently gather the breakdown characteristics that define this statistical process. The proposed circuit can monitor a progressive decrease in gate resistance, or simply an abrupt failure often referred to as a hard breakdown. This structure greatly reduces the required process characterization testing time, which may involve continuously monitoring the current through a single Device Under Test (DUT) per experiment with a finely tuned parametric test system. Given the need for up to thousands of test samples to correctly define the Weibull slope of the time to breakdown (T_{BD}) distribution [2], that serial testing process quickly becomes cumbersome. In addition, the array format is also a convenient method to study any spatial correlation of gate dielectric breakdown characteristics, without requiring sophisticated testers or elaborate test setups. Test array structures of this type are gaining popularity as a fast and efficient way to gather process technology information that is statistically meaningful, since individual device probing is not practical when large numbers of readings are required [6-7].

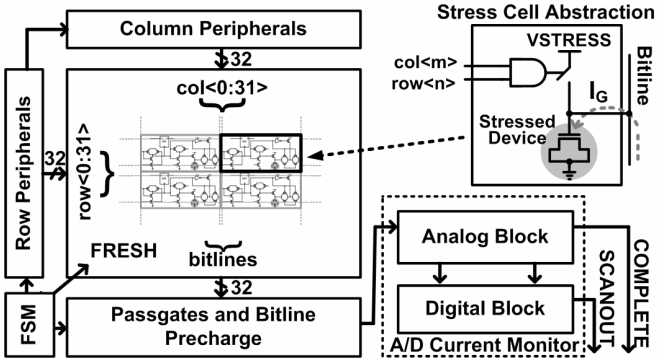


Fig. 1. 32x32 array for fully automated gate dielectric breakdown characterization.

II. Breakdown Characterization Array Design

The proposed test circuit design consists of a 32x32 array of stressed NMOS transistors, whose gate currents (I_G) are periodically measured using an A/D current monitor and on-chip control logic (Fig. 1). (Note that PMOS transistors could be tested within the same framework with the addition of a slightly modified stress cell design.) After an initialization sequence, cells are cycled through automatically without the need to send or decode cell addresses, in order to simplify the logic and attain faster measurement times. A single external clock signal is asserted each time that the controlling software is ready for a new I_G measurement. Although we chose to simplify and speed up the circuit in this manner, we do have the ability to select any one portion of the array for measurements while turning off stress in the rest of the test cells, as will be described later. The finite state machine (FSM) pictured in Fig. 1 controls the initialization sequence timing, as well as that of the subsequent measurements. The row and column peripheral circuits contain D flip-flops and multiplexers used to select a particular cell, as well as level shifters to boost signals from the 1.2V digital supply domain to the stress voltage (VSTRESS) level, which is used as the supply voltage within the array.

The stress cell structure shown in Fig. 2 was implemented to facilitate the accelerated stressing of the DUTs, by using thick oxide I/O transistors in the supporting circuitry to avoid excessive aging or breakdown in these other devices. The row<n> and col<m> signals are used to select one stress cell when both are logic high. When a cell is selected, devices M1 and M2 are turned off, and the transmission gates connecting the gate of the DUT to its bitline are turned on. M2 precharges the node between the two transmission gates to VCC, which matches the bitline precharge level, until this cell selection event. After the transmission gates are turned on, the gate current through the stressed DUT is measured using circuits in the A/D current monitor.

The FRESH signal is used to permanently gate off stress on a broken device when a high gate current is detected, in order to avoid excessive current drawn from the VSTRESS supply. After a sufficiently high breakdown current is measured in a selected cell, FRESH is automatically set to 0V by the controlling software before row<n> goes low, which latches a low value on Q. This isolates the DUT from VSTRESS by turning off device M1. When a cell is not being

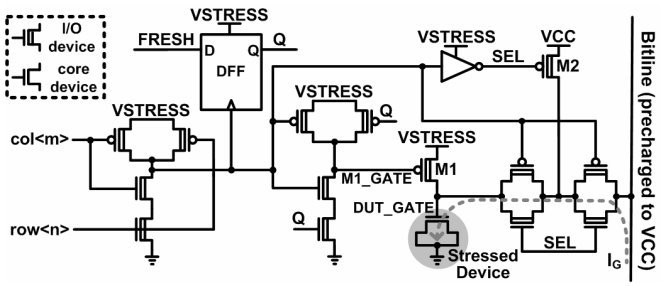


Fig. 2. Stress cell with bitline leakage compensation and stress/no-stress capability.

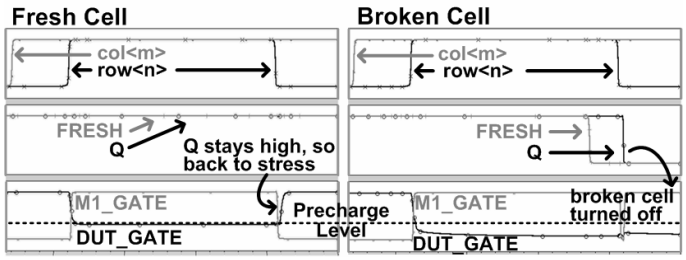


Fig. 3. Signal waveforms for one measurement in (a) a fresh cell and (b) a broken cell. The precharge level on the bitline (VCC) is indicated by the dashed line.

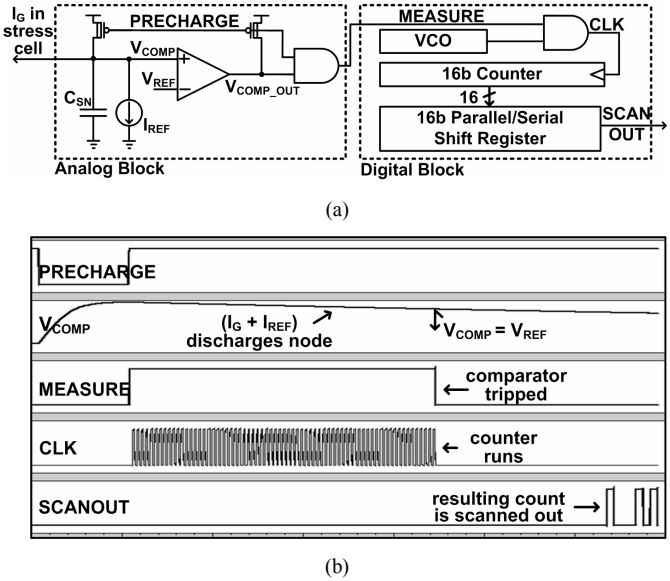


Fig. 4. (a) Analog and digital blocks of the A/D current monitor for measuring breakdown current progression. (b) Simulation of these blocks during a breakdown cell measurement

selected and Q is still high, M1 is turned on and the DUT is placed under constant voltage stress. Simulation waveforms demonstrating the measurement procedure for a fresh cell (low gate current) and a broken cell (high gate current) are presented in Fig. 3. In the unbroken, or "fresh" cell, the DUT_GATE node voltage quickly drops to the 1.2V precharge level before slowly decaying to ~V_{REF} (discussed later in this section), and then being charged to VSTRESS again when the row<n> signal drops to 0V. When the "broken" cell is accessed for a measurement the DUT_GATE node quickly discharges to 0V through the breakdown path. In this case, the FRESH signal is set to 0V before the cell is deselected, so M1_GATE remains high, and no further stressing occurs in this cell. The FRESH signal can also be used during circuit initialization to gate off stress in a range of unused cells which may be tested at a later time, or cells that are already broken from a previous experiment. This feature allows us to measure any one portion of the array during a single test, rather than the entire 1024 cells, if so desired.

The two transmission gates were placed between the gate of each DUT and its bitline, with the internal node held at VCC when the cell is not selected, to cut off the leakage in all

unselected stress cells from the A/D current monitor. Simulations show that the leakage sourced by all 1023 unselected cells in the array during a measurement is limited to roughly 108nA. This worst-case leakage on the discharge path occurs when the selected DUT's gate node has discharged to ~1.1V (the V_{REF} level) at 30ºC and VCC = 1.2V.

The analog block shown in Fig. 4(a) contains a comparator whose precharged output drops to 0V when the precharged input voltage (stored on an 80pF metal capacitor, C_{SN}) falls to the reference voltage (V_{REF}) level. That discharge rate is determined by I_G in the selected cell, plus an external reference current (I_{REF}) that is used for calibration purposes. The digital block (Fig. 4(a)) contains a 16 bit counter that runs at a rate set by a voltage controlled oscillator (VCO), starting at the end of the precharge event until the comparator's output falls, indicating that a measurement is complete. Therefore, less I_G (i.e., a larger gate resistance, R_{GATE}) translates to a higher count. The final result is latched into a parallel/serial shift register and subsequently scanned off-chip after the software interface detects a COMPLETE signal, which is asserted by the analog block. The results are stored in a convenient spreadsheet format for post-processing. The simulation waveforms presented in Fig. 4(b) illustrate the basic outline of this measurement procedure.

III. Test Chip Measurements

A test chip was fabricated in a 1.2V, 130nm process, and automatic measurements were completed via LabVIEW™. The calibration procedure and measurement results are illustrated in Fig. 5. In order to obtain the final count vs. total discharge path resistance characteristic, the A/D current monitor was gated off from the breakdown array, and an adjustable external resistor (R_{EXT}) was attached to the I_{REF} path. Each output count during measurements can be translated into a gate path resistance by using this calibration curve, and the relationship $R_{TOTAL} = R_{EXT} \parallel R_{GATE}$, with R_{EXT} being fixed during measurements.

The resistance of the transmission gates located on the path from V_{COMP} to the DUT gates is not accounted for in this simple procedure, and therefore introduces a measurement error that becomes more severe as the DUT gate resistance drops into the hard breakdown region. That is, our measured R_{GATE} result will be larger than the correct value because of the transmission gate resistances. However, due to the relatively high value of R_{GATE} during the progressive degradation stage leading up to the final hard breakdown, this error is small in the region of interest. The measurement error is less than 5% for R_{GATE} values of 64kΩ and greater, corresponding to gate currents up to 18.8μA at a sensing voltage of 1.2V. Several authors have indicated that the soft to progressive breakdown regimes are well within this current limit [4-5]. Also note that a more detailed calibration path could be included in future implementations to take the transmission gate resistances and other circuit parasitics into account during calibration.

As seen in the direct device probing results of Fig. 6, we did not typically observe progressive dielectric breakdown in the CMOS process used here during accelerated measurements with stress voltages of ≥4V, when recording four measurements per second. Therefore, although the proposed design is capable of monitoring progressive breakdowns, we were specifically looking for hard breakdowns in the automated array measurements presented here. These events were defined as a sudden decrease in the scanned out discharge time count of at least an order of magnitude.

Cumulative distribution functions (CDF) of the time to breakdown for a range of stress voltages, both on a standard percentage scale as well as the Weibull scale, are displayed in

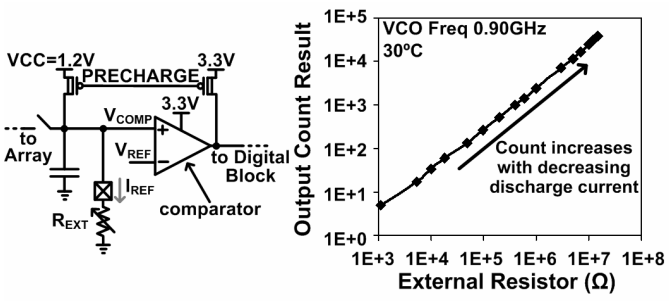


Fig. 5. Calibration setup and measured results.

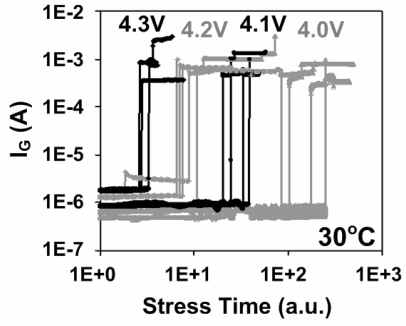
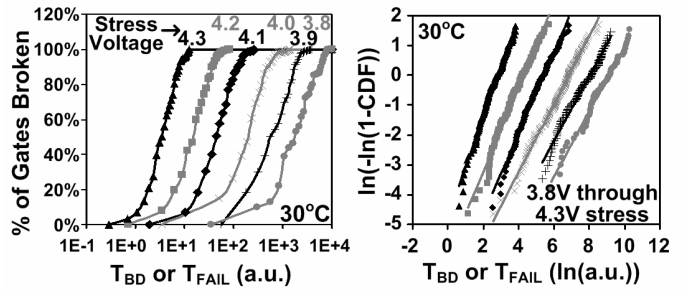
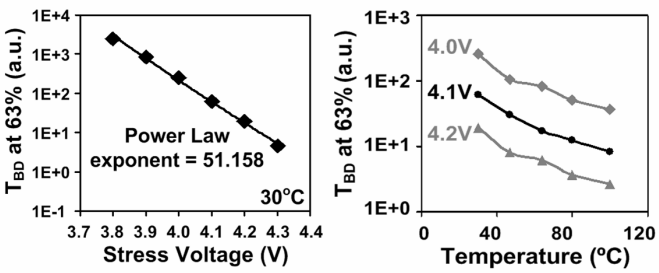


Fig. 6. Discrete device probing results showing hard breakdown behavior.



(a) (b)

Fig. 7. Measured T_{BD} CDF on (a) a percentage scale and (b) a Weibull scale.



(a) (b)

Fig. 8. (a) Voltage acceleration of T_{BD} at the 63% point. (b) TBD at the 63% point vs. temperature.

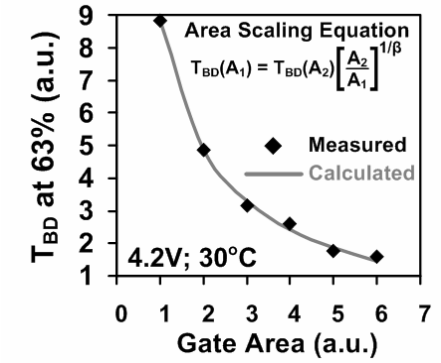


Fig. 9. Area scaling data, computed from measurement results using the weakest link characteristic of TDDB, compared with theoretical results [2].

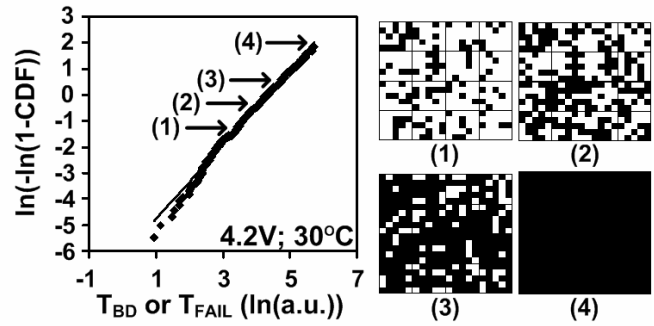
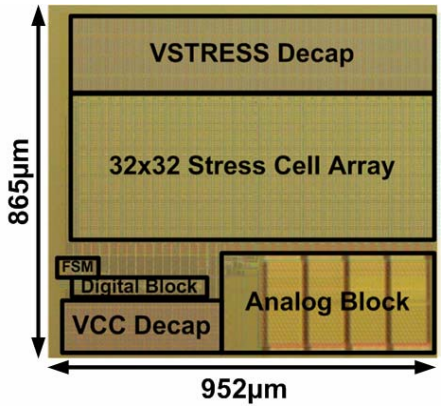


Fig. 10. Spatial T_{BD} distribution in a stress cell array at four time points on the Weibull scale CDF.

Fig. 7. The Weibull slope factor (β) for 4.2V stress was 1.443, with that factor slightly decreasing for lower stress voltages, and increasing at 4.3V. The exponential relationship of the Weibull characteristic life (time at which 63% of the devices have failed) with voltage is illustrated in Fig. 8(a). The power law exponent of this plot is ~51, which is slightly larger than that reported in previous work where the time to the first (soft) breakdown was recorded [8]. The dependency of the time to breakdown on stress temperature is shown in Fig. 8(b) for a range of voltages. Fig. 9 compares measured and calculated area scaling characteristics [2, 8]. The measured numbers were obtained by combining the results for a given number of spatially adjacent DUTs and selecting the smallest T_{BD} from each group, due to the weakest-link character of dielectric breakdown. In Fig. 10, the spatial distribution of T_{BD} in a 20x20 portion of a test array stressed at 4.2V is plotted along with the corresponding Weibull distribution. The four spatial diagrams correspond to the four divisions of the Weibull plot representing 25% of the cells each. As indicated in this figure, no spatial correlation was detected in these experiments. A test chip microphotograph and summary of the chip characteristics is shown in Fig. 11.

V. Conclusion

We have presented a circuit design for the efficient characterization of gate dielectric breakdown during process characterization. The proposed design consists of a large array of test cells that facilitate the accelerated stressing of the devices under test and an A/D current monitor that translates



Technology	0.13µm CMOS
Digital Supply	1.2V
Dimensions	952x865µm2
Gate Resisistance Measurement Range	~1kΩ - 10MΩ

Fig. 11. Microphotograph and summary of the test chip characteristics.

the gate current of each measured device into a digital count that is scanned off chip for post processing. A simple automated design such as this could greatly reduce testing times, as up to thousands of samples are needed to correctly define the statistical characteristics of TDDB. A range of test chip measurements from a 32x32 array implemented in a 1.2V, 130nm CMOS process were presented to demonstrate the functionality and flexibility of this design.

Acknowledgments

The authors would like to thank Intel and IBM for financial support, as well as United Microelectronics Corporation (UMC) for the foundry design kit and chip fabrication.

References

[1] R. Degraeve, et al., "New Insights in the Relation Between Electron Trap Generation and the Statistical Properties of Oxide Breakdown," Trans. on Electron Devices, vol. 45, no. 4, pp. 904-911, 1998.

[2] E. Y. Wu, et al., "CMOS scaling beyond the 100-nm node with Silicon-Dioxide-Based Gate Dielectrics," IBM J. of R & D, pp. 287-298, Vol. 46, No. 2/3, 2002.

[3] J. H. Stathis, "Gate Oxide Reliability for Nano-Scale CMOS," Int. Conf. on Microelectonics, pp. 78-83, 2006.

[4] J. Suñé, et al., "Statistics of Competing Post-Breakdown Failure Modes in Ultrathin MOS Devices," Trans. on Electron Devices, vol. 53, no. 2, pp. 224-234, 2006.

[5] A. Kerber, "Lifetime Prediction for CMOS Devices with Ultra Thin Gate Oxides Based on Progressive Breakdown," Int. Reliability Physics Symp., pp. 217-220, 2007.

[6] L. Pang and B. Nikolic, "Impact of Layout on 90nm CMOS Process Parameter Fluctuations," Symp. on VLSI Circuits, pp. 69-70, 2006.

[7] K. Agarwal, et al., "A Test Structure for Characterizing Local Device Mismatches," Symp. on VLSI Circuits, pp. 67-68, 2006.

[8] E. Y. Wu, et al., "Experimental Evidence of T_{BD} Power-Law for Voltage Dependence of Oxide Breakdown in Ultrathin Gate Oxides," Trans. On Electron Devices, vol. 49, no. 12, pp. 2244-2253, 2002.