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

John Keane, *Student Member, IEEE*, Shrinivas Venkatraman, Paulo Butzen, *Student Member, IEEE*, and Chris H. Kim, *Member, IEEE*

***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 Weibull 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 32 × 32 test array implemented in a 130-nm bulk CMOS process. Results show that this system is capable of taking accurate measurements across a range of voltages and temperatures, which is critical for extrapolating accelerated stress experiment results to expected device lifetimes under realistic operating conditions.**

***Index Terms*—Aging, circuit reliability, dielectric breakdown, digital measurements.**

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) in gate stacks. This mechanism causes a conductive 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 build 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. In addition, the time to breakdown (T_{BD}) distributions for thinner gate dielectrics have a larger statistical spread over time [1], [2]. This can lead to large errors when extrapolating accelerated stress experiment results to realistic operating conditions and low failure percentiles in order to make device reliability predictions.

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 breakdown (see Fig. 1) [1]. This model is now being extended and modified to deal with ultra-thin oxides [3], [4]. Some studies have used the time to the first breakdown event (defined as an increase in gate current to some predetermined level) to extrapolate predicted device lifetimes from accelerated stress experiments [2], [5]. 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 breakdowns are more correctly described as progressive in nature [6].

In addition, it has become apparent that transistors can continue to function in certain cases after one or more breakdowns [see Fig. 1(b)], and the progressive, post-breakdown current evolution must also be taken into consideration to obtain less pessimistic lifetime projections [6]–[8]. This is particularly true

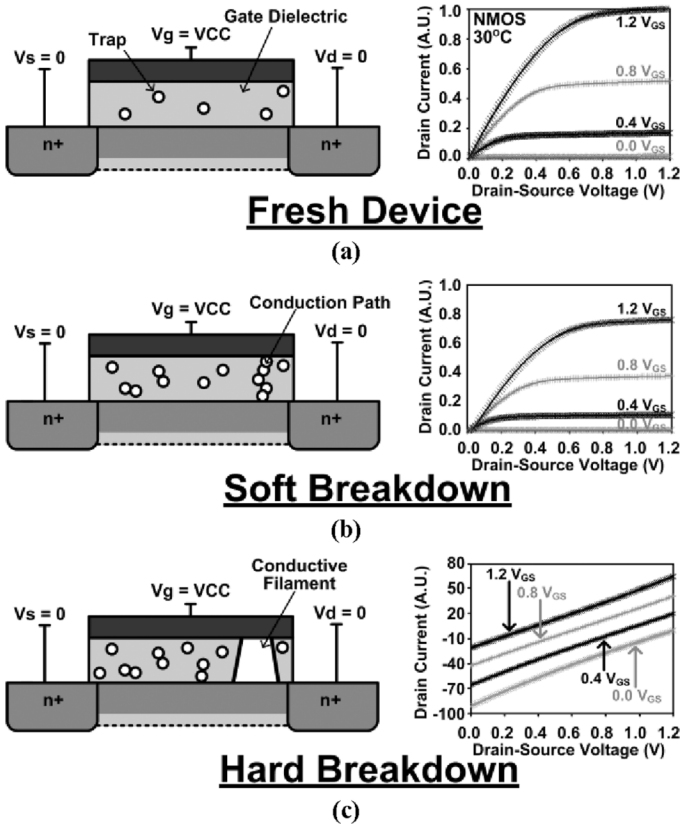

Fig. 1. Cross sections and measured I-V curves from device probing experiments on an nMOS device in a 130-nm bulk technology. These results illustrate the effects of a progressive soft-to-hard breakdown.

Manuscript received August 29, 2009; revised December 14, 2009. First published February 22, 2010; current version published April 27, 2011. This work was supported in part by the SRC under Award 2008-HJ-1805, along with Samsung, Intel, IBM, TI, and UMC.

J. Keane, S. Venkatraman, and C. H. Kim are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: jkeane@ece.umn.edu).

P. Butzen is with the Instituto de Informática-UFRGS, Federal University of Rio Grande do Sul, Porto Alegre 91.509–900, Brazil.

Digital Object Identifier 10.1109/TVLSI.2010.2041258

when operating at lower voltages and with thinner dielectrics, making an observable progressive breakdown current more likely before final device failure.

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 specific 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.

Most of the previously published TDDB measurement results were gathered from individual device probing experiments. The equipment used in those tests can be expensive, and testing each device individually leads to long experiment times. However, one on-chip circuits-based method to monitor gate dielectric wear-out was recently proposed [9]. In this case, results were provided in the form of a frequency shift of a Schmitt trigger oscillator which is modified by the increasing gate leakage through a pair of stressed PMOS devices. This design can provide some indication of the wear-out behavior of stressed transistors, but does not facilitate a direct reading of gate resistance degradation, or any other specific device characteristics (i.e., the end result is oscillator degradation with no suggestion of how to translate this into another parameter).

In this paper, we present a circuit design that performs automated measurements in a test array to directly 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 time, which may involve continuously monitoring the current through a single device under test (DUT) per experiment with a parametric test system. Given the need for up to thousands of samples to correctly define the Weibull slope of the T_{BD} distribution [2], [10], that serial testing process quickly becomes cumbersome. Therefore, in the circuit presented here, DUTs are stressed in parallel and we continuously loop through the array, temporarily removing stress conditions in one cell at a time and measuring each DUT's gate current. In addition, the array format is a convenient method to study any spatial correlation of TDDB without requiring elaborate test setups. Test array structures are gaining popularity as an efficient way to gather process technology information, since individual device probing is not convenient when large numbers of readings are required [11], [12].

II. Breakdown Characterization Array Design

The proposed test circuit design consists of a 32×32 array of structures we call "stress cells" that contain the DUTs, whose gate currents $(\mathrm{I_G})$ are periodically measured using an analog-to-digital (A/D) current monitor and on-chip control logic (see Fig. 2). 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

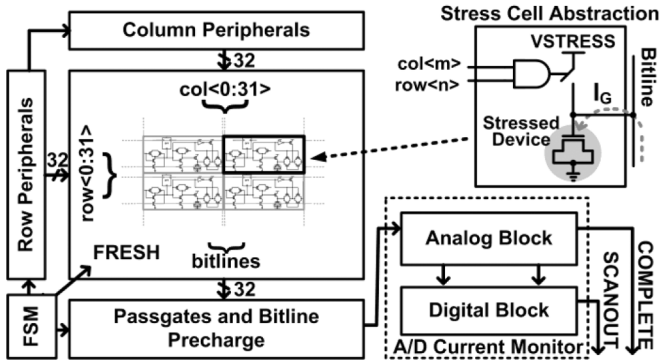

Fig. 2. Top level diagram of the 32×32 array for fully automated gate dielectric breakdown characterization.

faster measurement times. A single external clock signal is asserted each time that the controlling software is ready for a new $\mathrm{I_G}$ 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 discussed later. The finite state machine (FSM) in Fig. 2 controls the initialization sequence timing, as well as that of the subsequent measurements. The row and column peripheral circuits contain D flip-flops (DFFs) and multiplexers used to select a particular cell, as well as level shifters to boost signals from the 1.2 V (VCC) digital supply domain to the stress voltage (VSTRESS) level, which is used as the supply voltage within the array.

A. Stress Cell Design

The stress cell structure shown in Fig. 3(a) 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 dual-oxide requirement for the present design is commonly met by modern processes, but we currently have work underway to implement stressing circuits with a single oxide thickness.) Transistor M1 drives the DUT gate to VSTRESS if the cell has been turned on for a stress test. At the same time, M2 holds the node between the two pictured transmission gates at VCC, which matches the bitline precharge level, until the cell is selected for a measurement. The $\mathrm{row}\langle n\rangle$ and $\mathrm{col}\langle m\rangle$ signals are used to execute this selection event by setting both to the logic high level. At that time, devices M1 and M2 are turned off, and the transmission gates connecting the gate of the DUT to its bitline are turned on. After these steps are taken, the gate current through the stressed DUT is measured by the A/D current monitor.

The FRESH signal is used to permanently gate off stress on a broken DUT when a high gate current is detected, in order to avoid excessive current draw from the VSTRESS supply. After a sufficiently high breakdown current is measured in a selected cell, FRESH is set to 0 V by the controlling software before $\mathrm{row}\langle n\rangle$ goes low, which latches a logic low value on Q. This isolates the DUT from VSTRESS by turning off device M1. When a cell is not being selected for measurement and Q is still high, M1 is turned on and the DUT is placed under constant voltage stress. Note that the M1 devices should be sized to model a realistic gate driver. The current limitation of the driving stages in

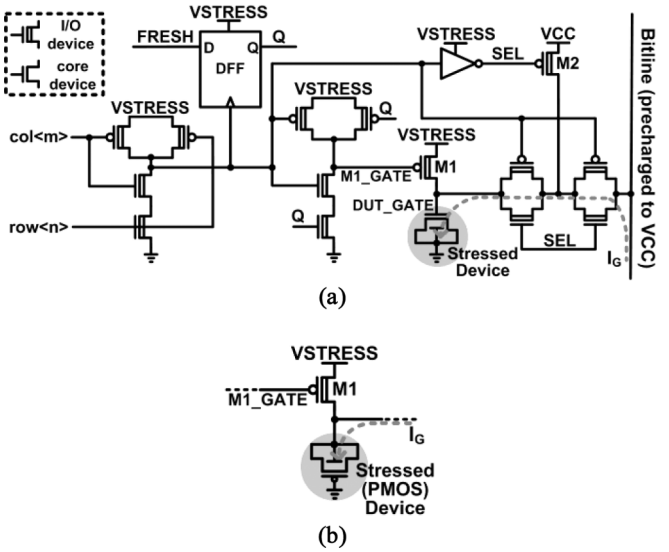

Fig. 3. (a) nMOS stress cell with bitline leakage compensation and stress/no-stress capability. (b) A pMOS stress cell would be identical to that seen in (a), with the change illustrated here. Note that the pMOS DUT requires its own isolated nwell.

TDDB experiments, such as M1 in this case, have been reported to strongly influence post-breakdown characteristics [13]. However, it was stated in that same work that the time until the first breakdown is not changed by the strength of these drivers.

pMOS transistors could be tested within this same framework by changing the DUT configuration as shown in Fig. 3(b). In this case, the pMOS DUT would be contained in an isolated nwell, and the drain and source would be connected to its body contact. The gate terminal would stay grounded, while the other three terminals are stressed or left floating for current measurements. In the first implementation presented here, we used only nMOS devices for simplicity and consistency.

Simulation waveforms demonstrating the measurement procedure for a fresh cell (i.e., DUT with low $\mathrm{I_G}$) and a broken cell (i.e., DUT with high $\mathrm{I_G}$) are presented in Fig. 4(a) and (b), respectively. In the unbroken, or "fresh" cell, the DUT_GATE node voltage drops to the 1.2 V precharge level before slowly decaying to $\sim V_{\mathrm{REF}}$ (a value defined in Section II-B), and then being charged to VSTRESS again when the $\mathrm{row}\langle n\rangle$ signal drops to 0 V. When the "broken" cell is accessed for a measurement the DUT_GATE node discharges to 0 V through the breakdown path. In this case, the FRESH signal is set to 0 V before the cell is deselected, so M1_GATE remains high, and no further stressing occurs in this cell. We do not expect that the voltage transients on the DUT_GATE node during the measurement transitions should significantly impact the breakdown process since no voltage overshoots are observed. Also, several reports have found that time to breakdown simply increases with shorter stress duty cycle, rather than being negatively impacted by the switching activity [14], [15]. Finally, the drop from VSTRESS to VCC on the DUT_GATE does not impact the gate resistance measurement because the bitline is held at VCC by the precharge device until after the cell is selected.

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

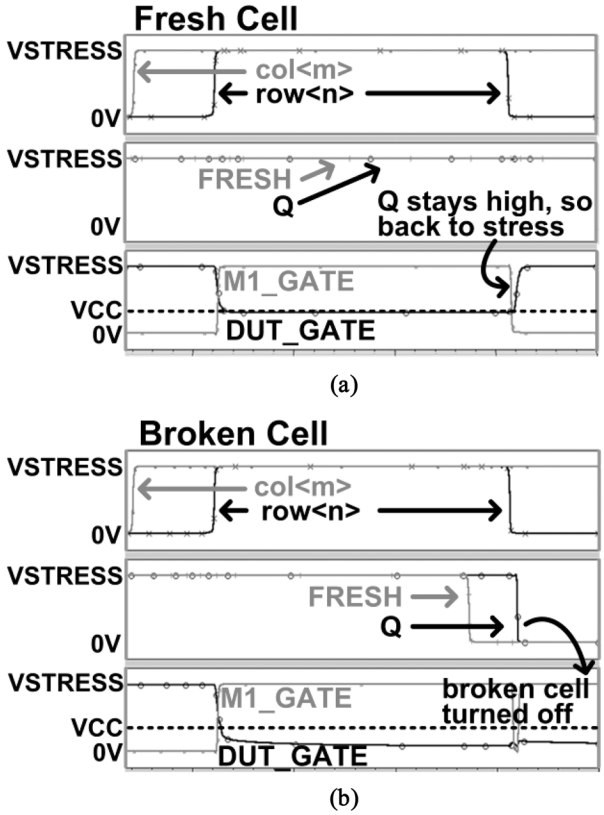

Fig. 4. Simulated stress cell operation corresponding to Fig. 3(a). (a) Illustrates a measurement taking place in a fresh (i.e., prebreakdown) cell. (b) Illustrates a measurement in a broken cell.

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. In the cell range selection step, we leave VSTRESS at 1.2 V during the first loop through the entire array, while setting FRESH low for those cells that we do not wish to measure during subsequent loops. The thick oxide I/O transistors in the stress cells operate correctly at this low voltage level, and we expect that no appreciable gate dielectric degradation will occur in the DUTs at their nominal supply voltage over the course of a few seconds. After this initialization loop, VSTRESS is raised to the stressing voltage, and measurements proceed as usual in the selected portion of the array.

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, in order to keep leakage between all unselected stress cells and the A/D current monitor low and consistent. Simulations show that the total leakage sourced by all 1023 unselected cells in the array during a measurement is limited to $\sim$ 108 nA. This worst-case leakage on the discharge path occurs when the selected DUT's gate node has discharged to $\sim$ 1.1 V (the V_{REF} level) at 30 °C and $\mathrm{VCC} = 1.2$ V.

B. A/D Current Monitor

The analog block shown in Fig. 5(a) contains a comparator whose precharged output drops to 0 V when the precharged input voltage (stored on an 80 pF metal capacitor, C_{SN}) falls below the reference voltage (V_{REF}) level. That discharge rate is determined by $\mathrm{I_G}$ in the selected cell, plus an external reference

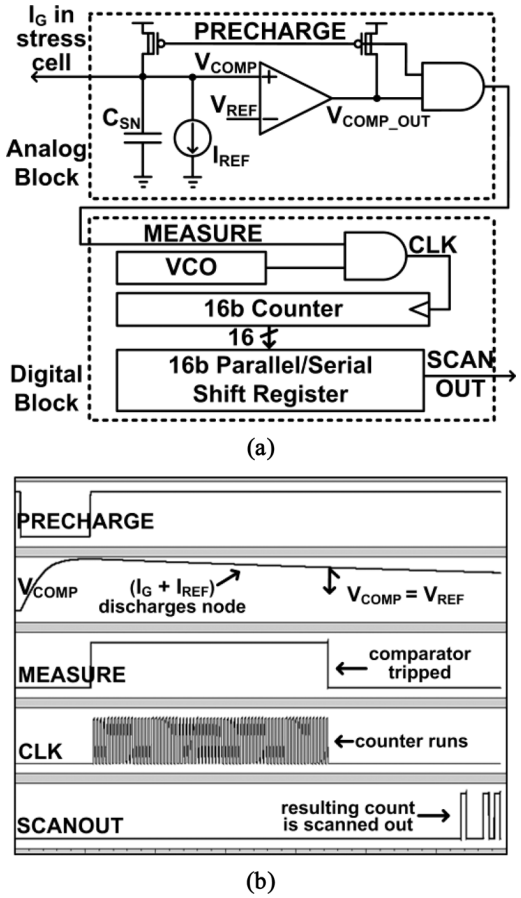

Fig. 5. (a) A/D current monitor used to translate the gate current through a DUT $(\mathrm{I_G})$ into a 16 bit digital count. (b) Simulation of this A/D conversion.

current $(\mathrm{I_{REF}})$ that is also used for calibration purposes. The digital block [see Fig. 5(a)] contains a 16 bit counter that runs at a rate set by a voltage controlled oscillator (VCO), from the end of the precharge event until the comparator's output falls, indicating that a measurement is complete. Therefore, lower $\mathrm{I_G}$ (i.e., a larger gate resistance, $\mathrm{R_{GATE}}$) translates into a higher count, and vice versa.

The final count result is latched into a parallel/serial shift register and scanned off-chip after the software interface detects a completion signal, which is asserted by the analog block. The results are stored in a convenient spreadsheet format for post-processing. A calibration technique described in the Section III-A makes it possible to translate these resulting counts into gate resistance values, so we can monitor a progressive breakdown process in each of the stressed devices by running measurements in a continuous loop through the array. The simulation waveforms presented in Fig. 5(b) illustrate the basic outline of this measurement procedure.

C. Peripheral Circuits and Operational Flow

The first two rows of the row peripheral block from Fig. 2 are illustrated in Fig. 6(a). These circuits are identical to those in the column peripherals, but the latter are only clocked once after each time an entire column of cells has been selected individually. As mentioned earlier, cells are cycled through automatically at each pulse of the internal clock (Φ_{INT}) without

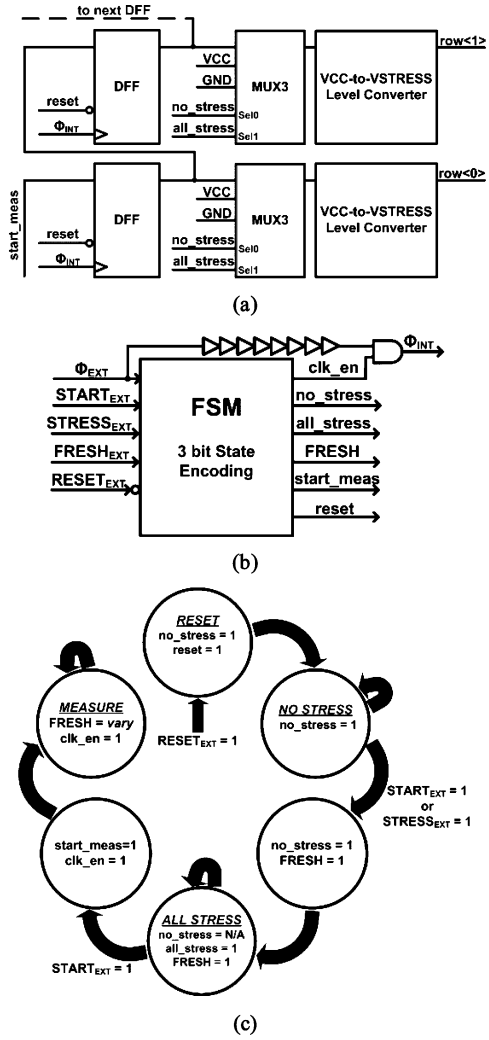

Fig. 6. (a) Block diagram of the first two rows of the row peripherals from Fig. 2. The column peripherals are identical to this, but are only clocked once after each time an entire column of cells has been accessed. (b) I/O diagram of the FSM which uses three bit state encoding. (c) State transition diagram. Transitions occur with each assertion of the external clock signal (Φ_{EXT}). Note that all internal signals not show explicitly in each state are set to 0 V.

the need to send or decode cell addresses, in order to simplify the logic and attain faster measurement times. The all_stress and no_stress select signals for the three-way MUX are used during circuit initialization, or to hold the array in a steady state where either all cells or no cells are stressed.

Fig. 6(b) shows an I/O diagram for the finite-state machine (FSM), which uses three bit state encoding, and the corresponding state transition diagram is presented in Fig. 6(c). State transitions, or moves to the next cell to be tested during the *MEASURE* state, occur with the assertion of an external clock signal (Φ_{EXT}) and depend on the current state and FSM inputs. Note that internal signals not shown explicitly in this transition diagram are set to 0 V. When the external reset signal $(\mathrm{RESET_{EXT}})$ is asserted at any point during operation, the measurement system enters the *RESET* stage, where all DFF outputs are driven to 0 V and stress is turned off in all cells. The latter is accomplished by setting the no_stress signal high, which drives all peripheral MUX3 outputs to VCC, thereby

selecting all cells to turn off all M1 transistors [see Fig. 3(a)]. After the next assertion of $\Phi_{\rm EXT}$, the unstressed array will wait in the *NO STRESS* state until $\rm STRESS_{EXT}$ is set to logic high, meaning that we wish to enter a constant stressing state for all cells (*ALL STRESS*), or $\rm START_{EXT}$ is asserted indicating that we want to start normal stress/measurement operation (*MEASURE*). In either case, the $\rm FRESH_{EXT}$ signal is first set high before deselecting all cells in order to clock all of the DFFs within the stress cells [see Fig. 3(a)], and set all Q values in those cells high. All cells are then deselected by setting all_stress high, which drives all peripheral MUX3 outputs to 0 V. If $\rm START_{EXT}$ is asserted, we continue from this step into normal stress/measurement operation.

As stated earlier, VSTRESS is left at the nominal digital supply voltage of 1.2 V during the short circuit initialization period in order to prevent accelerated stressing of the DUT gates. It is then held at 1.2 V throughout the first measurement loop if we wish to keep only a selected portion of the cells on for stress by appropriately asserting the $\rm FRESH_{EXT}$ signal. Immediately after this startup phase, VSTRESS is raised to the stressing voltage, and measurements proceed as usual in the selected portion of the array.

The on-chip phase of the measurement after each cell selection event required $\sim$100 μs or less, as determined by the values of $\rm C_{SN}, I_{REF}, I_G$, and additional leakage currents. However, the timing bottleneck was the results scanout routine executed by the controlling software. This portion of each measurement led to a total measurement time of several hundred milliseconds. Therefore, in order to keep a reasonable timing resolution of roughly 15 s or less between sequential checks of each stressed cell (depending on the VSTRESS value), we limited the number of cells tested in any one run. For example, experiments often covered a 5×5 portion of the array. As explained earlier, the FRESH signal in each cell was set such that the stress cells not used during any particular experiment were turned off. In the future, an improved software interface or different data acquisition board could be used to greatly reduce the results scanout time.

III. Test Chip Measurements

A test chip was fabricated in a 1.2 V, 130-nm bulk CMOS process. Each nMOS device under test had a width and length of 2 μm. Information about the gate dielectric construction is confidential, but the thickness is within a reasonable range for this technology node. Automatic measurements were completed with LabVIEW software and a National Instruments data acquisition board, which was connected to a laptop through a USB port. A microphotograph of the chip and a summary of the circuit characteristics are shown in Fig. 7. In the picture on the right of this figure, which captures a larger portion of the total chip, we point out a number of individual devices that were fabricated to verify that the results from our array match well with probing measurements. The probing experiments were completed with an HP semiconductor parameter analyzer and a Signature probe station.

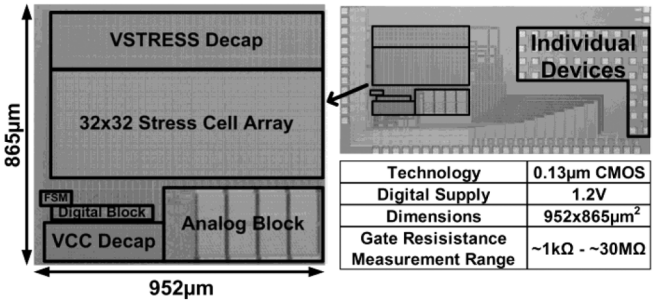

Fig. 7. Microphotograph and summary of the test chip characteristics. The individual devices reserved for probing experiments are labeled to the right of the TDDB array measurement system.

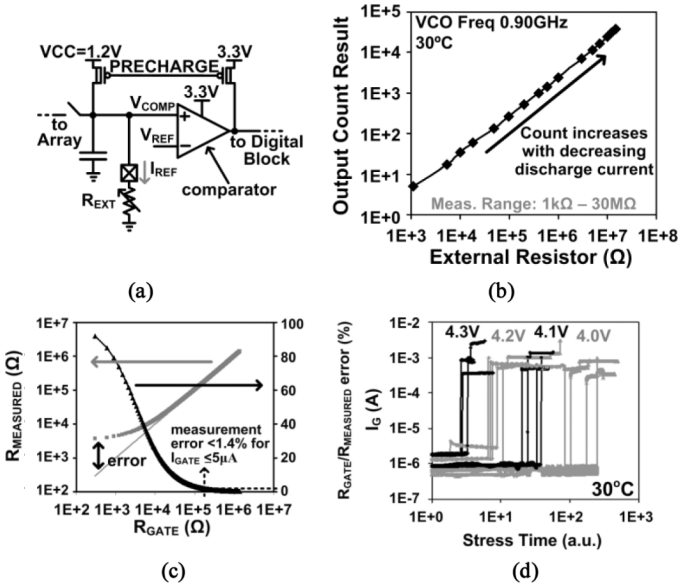

Fig. 8. (a) Measurement calibration setup. (b) Measured calibration results. (c) The resistance of the transmission gates located on the path from $V_{\rm COMP}$ to the DUT gates is not accounted for in this calibration procedure, but only introduces a small measurement error in the progressive breakdown region. (d) Individual device probing results indicate that in the stress voltage range of interest, with a sampling rate of 4 Hz, we expect to observe hard breakdowns in the majority of our experiments.

A. Test Chip Calibration Procedure

The calibration procedure and measurement results are illustrated in Fig. 8(a) and (b), respectively. In order to obtain the final count versus total discharge path resistance characteristic, the A/D current monitor was gated off from the breakdown array, and an adjustable external resistor $(\rm R_{EXT})$ was attached to the $\rm I_{REF}$ path. Therefore the total discharge path resistance $(\rm R_{TOTAL})$ in this case is simply the value of the external resistor. During this calibration procedure, the A/D current monitor is run normally, as it would during stress measurements, but with a range of known $\rm R_{TOTAL}$ values. This leads to a calibration curve, like that shown in Fig. 8(b).

After the calibration is completed, each output count recorded during stress measurements can be translated into a gate path resistance by using the calibration curve, and the simple equation $R_{\rm TOTAL} = R_{\rm EXT} \| R_{\rm GATE}$. Throughout measurements, $R_{\rm EXT}$ is held at a known constant value, and $R_{\rm TOTAL}$ is taken from the calibration curve at the point with an equivalent output count

result (i.e., the stress measurement count result matches a certain calibration count result), so R_{GATE} is the only unknown.

The range of gate resistances that this array-based system is able to record is roughly bounded from above by the value of R_{EXT}, since the smaller value of $R_{\mathrm{EXT}}||R_{\mathrm{GATE}}$ will dominate this equation, as well as the size of the counter that the VCO clocks during measurements. As explained in Section II-B, a larger R_{GATE} leads to longer C_{SN} discharge times, and hence higher count results. Therefore, measuring high values of R_{GATE} requires a sufficiently large counter. The lower bound of the measurement range is set by the speed of this VCO, because a higher clock rate is required to maintain sufficient resolution with faster discharge times. These bounds should be appropriately adjusted at design time, as well as during calibration. As we show in Fig. 8(b), the present design achieves an R_{GATE} measurement range of $\sim$ 1 kΩ–$\sim$ 30 MΩ (following the plotted trend through a count of 2^{16}) with the VCO clocked at 900 MHz.

The resistance of the transmission gates located on the path from V_{COMP} to the DUT gates is not accounted for in this calibration procedure, and therefore introduces error that becomes more severe as the DUT gate resistance drops into the hard breakdown region. That is, our measured R_{GATE} results will be larger than the correct value because of the additional 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, as shown in Fig. 8(c). The measurement error is less than 1.4% for R_{GATE} values of 240 kΩ and greater, corresponding to gate currents up to 5 μA at a sensing voltage of 1.2 V. Several authors have indicated that the soft to progressive breakdown regimes are within this current limit [7], [8].

A more detailed calibration path that duplicates the additional transmission gate resistances and other non-idealities could be included in future test chips to eliminate this small error. For example, the circuit could include replica cells embedded within the measurement array for calibration. An external resistor could then be attached directly to the node within those cells where a DUT gate would regularly be located. This procedure would exactly duplicate the normal measurement routine so that all leakages and parasitics are accounted for.

However, as seen in the direct device probing results of Fig. 8(d), we typically did not observe progressive dielectric breakdown in the CMOS process used here. This data was recorded during accelerated measurements with stress voltages of 4 V, 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 our automated array measurements. These events were defined as a sudden and sustained decrease in the scanned out discharge time count of roughly two orders of magnitude, when the VCO clocking the counter in the A/D current monitor was running at 900 MHz.

B. Measured Breakdown Distributions

Cumulative distribution functions (CDF) of the time to breakdown, both on a standard percentage scale as well as the Weibull scale, are displayed in Fig. 9(a) and (b), respectively. That data

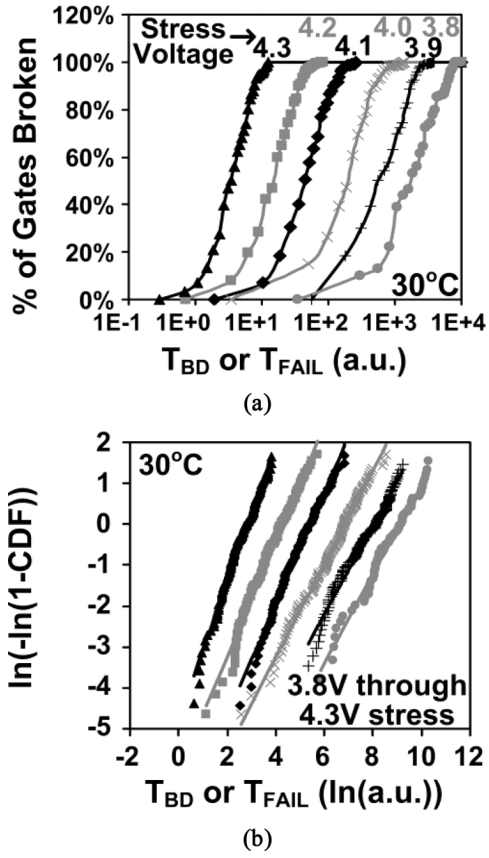

Fig. 9. Measured T_{BD} CDFs on (a) a standard percentage scale and (b) a Weibull scale.

was gathered at 30 °C, with stress voltages from 3.8 to 4.3 V in this 1.2 V process. TDDB follows Weibull statistics because this mechanism has a weakest-link character, since there are a large number of spots in each gate where the first breakdown can occur, and the breakdown process proceeds independently at each of them. The first breakdown at any of those locations leads to device degradation or failure though, so it can be thought of as the "weakest link." When we have a weakest-link process, extreme value distributions are the first functions we try to fit to measured data. Since in the case of time to breakdown the distribution is bounded from below at time zero, specifically we use a Weibull distribution [16].

The Weibull slope factor (β) for 4.2 V stress was 1.443, with that factor slightly decreasing for lower stress voltages, and increasing at 4.3 V. We generally expect that the Weibull slope should be dependent on gate dielectric thickness, but not voltage, so this slight difference was not expected. However, the slope values are still in good agreement with other published data [2], [13]. This trend is also observed to some degree in the results presented by Röhner, although not mentioned explicitly [13]. Finally, in a recent publication by Tous exploring breakdown in ultra-thin gate oxides, an explanation was provided for steeper distribution slopes at lower ranges of T_{FAIL} on the Weibull plot [10]. This phenomenon was attributed to the non-Weibull shape of T_{FAIL} for very thin oxides, which is only correctly observed with sufficiently large test sample sizes (well over 100). For these reasons, the small variation in our measured

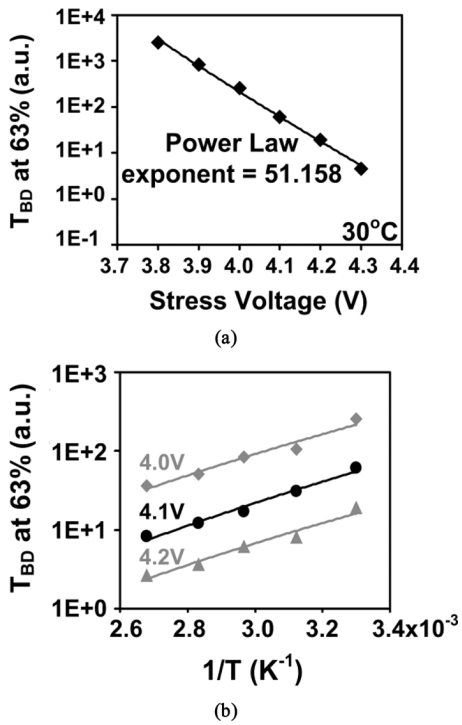

Fig. 10. (a) Voltage acceleration of $T_{\rm BD}$ at the 63% point. (b) $T_{\rm BD}$ at the 63% point versus the inverse of the temperature in Kelvins.

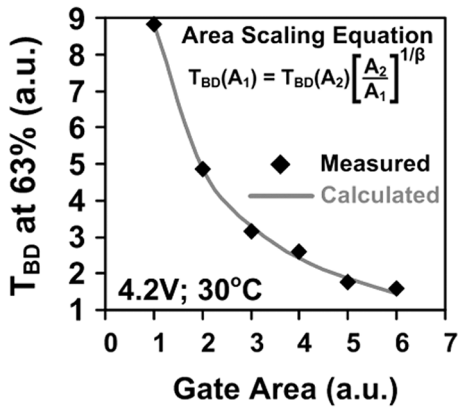

Fig. 11. Area scaling data computed from the combined measurement results of spatially adjacent stress cells, compared with theoretical results [2].

breakdown distribution slopes seems reasonable, and may have a theoretical justification.

C. Voltage and Temperature Acceleration of TDDB

The exponential relationship of the Weibull characteristic life (time at which 63% of the devices have failed) with voltage is illustrated in Fig. 10(a). The power law exponent is ~ 51, which is slightly larger than that reported in previous work where the time to the first breakdown event (soft or hard) was recorded [17]. Note that Wu *et al.* provided a physics-based explanation for voltage acceleration power law factors in the 40–50 range in that paper.

The measured dependency of the time to breakdown on stress temperature is shown in Fig. 10(b) for a range of voltages. In this temperature range of 30 °C to 100 °C, TDDB follows Arrhenius behavior with only small errors. Although the temperature dependence of breakdown is often modeled using Arrhenius behavior, non-Arrhenius dependence has also been reported at temperatures over 100 °C, particularly for thin gate dielectrics [18], [19]. At any rate, the temperature acceleration of TDDB imposes more severe limits on modern CMOS designs where device density and high clocking rates lead to increased local heating.

D. Area Scaling Property of TDDB

In addition to showing that the CDFs form straight lines on a Weibull scale in order to justify the use of these statistics to describe a process, we can also check that the process follows the unique area scaling property of this extreme value distribution (equation in Fig. 11). (Although it has long been established that TDDB follows Weibull statistics, we address this issue to illustrate the important concept of area scaling.) In our case since all DUTs are the same size, the measured numbers for different areas were obtained by combining the results for a given number of spatially adjacent DUTs. We then selected the smallest time to breakdown from each group, due to the weakest-link character of dielectric breakdown. The results shown in Fig. 11 indicate that our measured data matches well with the theoretical area scaling equation [2], [16], [17]. The scaling property is also used in other studies to define the Weibull slope parameter with a high degree of accuracy. That is done by measuring the time to breakdown for devices with a large area ratio, and then using the equation shown in Fig. 11 where the only unknown is β.

E. Spatial Distribution of Time to Breakdown

Test arrays such as ours, where a large number of devices are closely spaced, facilitate investigations of any spatial correlation in the process or characteristics being studied. For example, spatial correlation of gate oxide thicknesses could lead to a correspondingly correlated breakdown process [20]. The spatial distribution of $T_{\rm BD}$ in a 20×20 portion of a test array stressed at 4.2 V is plotted in Fig. 12, 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. No spatial correlation is apparent from these plots, and it is possible to check our conclusion with a quantitative measure of that phenomenon by calculating the local and global Moran's I statistics [21], [22]. However, this method works under the null hypothesis that the input data are normally distributed random variables, which we have seen is not the case for $T_{\rm BD}$ distributions. This is made clear in Fig. 13(a), where we plot the histogram of the data used to create Fig. 12.

The null hypothesis described above is common in statistical data analysis tools, so mathematicians have developed a number of methods to transform non-normal distributions to the normal form. The equation for the Box-Cox transformation, which can be used to transform Weibull distributions for this purpose, is shown in Fig. 13(b) [23]. This operation is defined by the λ in that equation, which in our case was calculated with the MATLAB "boxcox" function. The exact value found was 0.2833, and the resulting histogram is shown along with the transform equation. We verified the symmetry of this new data set with a "triples test" [24], [25]. The transformed data is also shown in

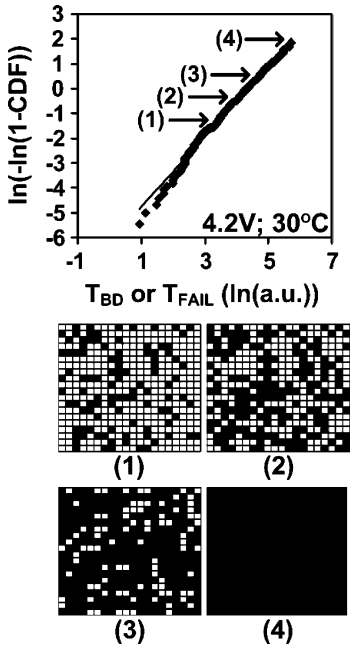

Fig. 12. Spatial distribution of TBD in a 20×20 stress cell array at four time points on the Weibull scale CDF. Cell locations are filled in once their DUT gates have broken down.

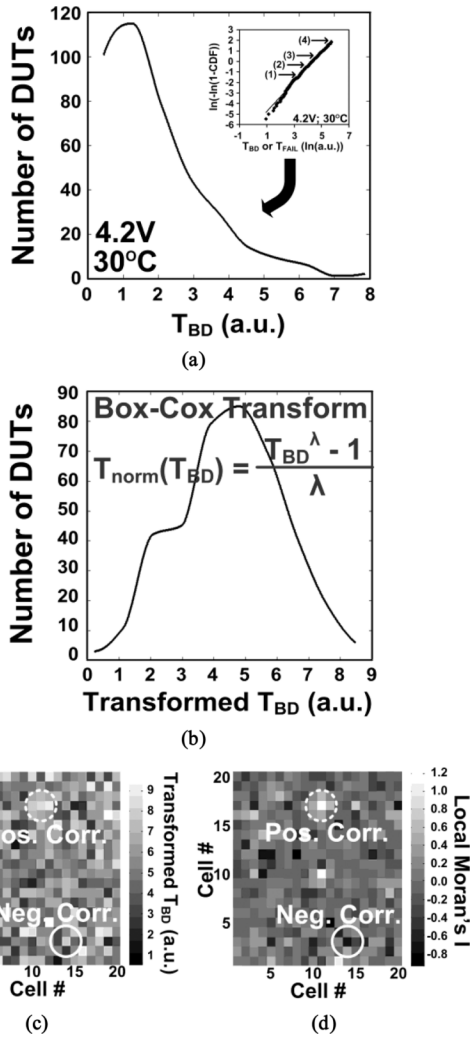

Fig. 13. (a) Histogram of time to breakdown in a 20×20 portion of a test array stressed at 4.2 V along with the corresponding Weibull plot from Fig. 12 (inset). (b) Histogram after the Box-Cox transformation is applied to create a normal distribution of $T_{\rm BD}$ data ($\lambda = 0.2833$). (c) Spatial diagram of the 20×20 array of cells with colors indicating each location's transformed $T_{\rm BD}$ [in arbitrary units matching those in part (b)]. (d) Local Moran's I for each cell location. Light colors in this last plot indicate positive correlation (i.e., "clustering") while darker colors indicate negative correlation (i.e., "dispersion").

a spatial plot in Fig. 13(c) with arbitrary units, matching those in Fig. 13(b). The area of this 20×20 array is roughly 555 μm $\times$ 225 μm in the physical implementation.

A sliding 3×3 contiguity matrix in the queen configuration was used to calculate the local Moran's I statistics [21]. This matrix defines the neighborhood around each value that is used to calculate spatial correlation, and the "queen" term is an analogy to chess. In this case, correlation with all eight nearest neighbors surrounding one cell is computed, and the results are plotted in Fig. 13(d). Lighter colors in this last plot indicate stronger positive correlation (i.e., "clustering") while darker colors indicate negative correlation (i.e., "dispersion"). Examples of both extremes are indicated. It is apparent that positive spatial correlation corresponds to cell locations in Fig. 13(c) that are surrounded by similar $T_{\rm BD}$ values, or similar colors in this plot format. The opposite is true for negative correlation. No strong correlation trend is observed, and the global Moran's I for this example was -8.907e-4, indicating negligible spatial correlation of $T_{\rm BD}$. No significant difference is observed in the results when a larger contiguity matrix is used.

IV. Conclusion

We have presented a circuit design for the efficient characterization of gate dielectric breakdown. The proposed system consists of a large array of test cells that facilitate the accelerated stressing of the DUTs without significant aging or breakdowns in the supporting circuitry. An A/D current monitor translates the gate current of each device into a convenient 16 bit digital count that is scanned off chip for post processing. Although in the technology used here, we generally only observed hard breakdowns, this design is capable of tracking a progressive decrease in a gate resistance with a high degree of accuracy down to the start of the hard breakdown region. Our automated array-based design would greatly reduce testing times, as up to thousands of samples are needed to correctly define the statistical characteristics of TDDB. Specifically, when compared with individual device probing, our proposed system can cut the test time down by a factor proportional to the number of devices under test, since all of these transistors are stressed in parallel in our circuit. A range of test chip measurements from a 32×32 array implemented in a 1.2 V, 130 nm bulk CMOS process were presented to demonstrate the functionality and flexibility of this design.

Acknowledgment

The authors would like to thank Samsung, Intel, IBM, TI and UMC for the technical feedback and chip fabrication.

REFERENCES

[1] R. Degraeve, G. Groeseneken, R. Bellens, J. Ogier, M. Depas, P. Roussel, and H. Maes, "New insights in the relation between electron trap generation and the statistical properties of oxide breakdown," *IEEE Trans. Electron. Devices*, vol. 45, no. 4, pp. 904–911, Apr. 1998.

[2] E. Wu, E. Nowak, A. Vayshenker, W. Lai, and D. Harmon, "CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics," *IBM J. R&D*, vol. 46, no. 2/3, pp. 287–298, 2002.

[3] J. Suñé, E. Wu, and S. Tous, "A physics-based deconstruction of the percolation model of oxide breakdown," *Microelectron. Eng.*, vol. 84, no. 9–10, pp. 1917–1920, 2007.

[4] A. Krishnan and P. Nicollian, "Analytical extension of the cell-based oxide breakdown model to full percolation and its implications," in *Proc. IEEE Int. Reliab. Phys. Symp.*, 2007, pp. 232–239.

[5] Y. Lee, N. Mielke, M. Agostinelli, S. Gupta, R. Lu, and W. McMahon, "Prediction of logic product failure due to thin-gate oxide breakdown," in *Proc. IEEE Int. Reliab. Phys. Symp.*, 2006, pp. 18–28.

[6] J. Stathis, "Gate oxide reliability for nano-scale CMOS," in *Proc. IEEE Int. Conf. Microelecton.*, 2006, pp. 78–83.

[7] J. Suñé, E. Wu, and W. Lai, "Statistics of competing post-breakdown failure modes in ultrathin MOS devices," *IEEE Trans. Electron. Devices*, vol. 53, no. 2, pp. 224–234, Feb. 2006.

[8] A. Kerber, "Lifetime prediction for CMOS devices with ultra thin gate oxides based on progressive breakdown," in *Proc. IEEE Int. Reliab. Phys. Symp.*, 2007, pp. 217–220.

[9] E. Karl, P. Singh, D. Blaauw, and D. Sylvester, "Compact in-situ sensors for monitoring negative-bias-temperature-instability effect and oxide degradation," in *Proc. IEEE Int. Solid State Circuits Conf.*, 2008, pp. 410–411.

[10] S. Tous, E. Wu, and J. Suñé, "A compact model for oxide breakdown failure distribution in ultrathin oxides showing progressive breakdown," *IEEE Electron. Device Lett.*, vol. 29, no. 8, pp. 949–951, Aug. 2008.

[11] L. Pang and B. Nikolic, "Impact of layout on 90 nm CMOS process parameter fluctuations," in *Proc. IEEE Symp. VLSI Circuits*, 2006, pp. 69–70.

[12] K. Agarwal, F. Liu, C. McDowell, S. Nassif, K. Nowka, M. Palmer, D. Acharyya, and J. Plusquellic, "A test structure for characterizing local device mismatches," in *Proc. IEEE Symp. VLSI Circuits*, 2006, pp. 67–68.

[13] M. Röhner, A. Kerber, and M. Kerber, "Voltage acceleration of TBD and its correlation to post breakdown conductivity of N- and P-channel MOSFETs," in *Proc. IEEE Int. Reliab. Phys. Symp.*, 2006, pp. 76–81.

[14] M. Nafria, D. Yelamos, J. Suñé, and X. Aymerich, "Frequency dependence of degradation and breakdown of thin SiO_2 films," *Quality Reliab. Eng. Int.*, vol. 11, no. 4, pp. 257–261, 1995.

[15] E. Rosenbaum and C. Hu, "High-frequency time-dependent breakdown of $Si0_2$," *IEEE Electron. Device Lett.*, vol. 12, no. 6, pp. 267–269, Jun. 1991.

[16] D. Wolters and J. Verwey, "Breakdown and Wear-Out Phenomena in SiO_2 Films," in *Instabilities in Silicon Devices*. Amsterdam, The Netherlands: Elsevier, 1986, ch. 6.

[17] E. Wu, A. Vayshenker, E. Nowak, J. Suñé, R. Vollertsen, W. Lai, and D. Harmon, "Experimental evidence of T_{BD} power-law for voltage dependence of oxide breakdown in ultrathin gate oxides," *IEEE Trans. Electron. Devices*, vol. 49, no. 12, pp. 2244–2253, Dec. 2002.

[18] D. DiMaria and J. Stathis, "Non-arrhenius temperature dependence of reliability in ultrathin silicon dielectric films," *Appl. Phys. Lett.*, vol. 74, no. 12, pp. 1752–, 1999.

[19] B. Kaczer, R. Degraeve, N. Pangon, and G. Groeseneken, "The influence of elevated temperature on degradation and lifetime prediction of thin silicon-dioxide films," *IEEE Trans. Electron. Devices*, vol. 47, no. 7, pp. 1514–1521, Jul. 2000.

[20] K. Chopra, C. Zhuo, D. Blaauw, and D. Sylvester, "A statistical approach for full-chip gate-oxide reliability analysis," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, 2008, pp. 698–705.

[21] L. Anselin, "Local indicators of spatial association—LISA," *Geographical Anal.*, vol. 27, no. 2, pp. 93–115, 1995.

[22] F. Hebeler, "Moran's I," June 20, 2007. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/13663

[23] G. Box and D. Cox, "An analysis of transformations," *J. Royal Statistical Soc.*, vol. 26, no. 2, pp. 211–252, 1964, Series B (Methodological).

[24] R. Randles, M. Flinger, G. Policello, and D. Wolfe, "An asymptotically distribution-free test for symmetry versus asymmetry," *J. Amer. Statistical Association*, vol. 75, no. 369, pp. 168–172, 1980.

[25] J. van der Geest, "Triplestest," May 8, 2008. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/19547

[26] J. Keane, S. Venkatraman, P. Butzen, and C. H. Kim, "An array-based test circuit for fully automated gate dielectric breakdown characterization," in *Proc. IEEE Custom Integr. Circuits Conf.*, 2008, pp. 121–124.

John Keane (SM'06) received the B.S. degree (*summa cum laude*) in computer engineering from the University of Notre Dame, Notre Dame, IN, in 2003 and the M.S. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2005, where he is currently pursuing the Ph.D. degree.

He has completed three internships with the IBM Research Lab in Austin, TX, and several more with other high tech companies. He will join Intel Corporation, Hillsboro, OR, in the spring of 2010. His research involves developing methods to monitor aging and variation mechanisms in advanced CMOS technologies, as well as low power design issues.

Mr. Keane was a recipient of the University of Minnesota Graduate School Fellowship for the 2003–2005 academic years, along with IBM Ph.D. Fellowships in 2008 and 2009. In 2009, he was selected as an award winner in the DAC/ISSCC Student Design Contest. He won Best Paper in Session at the 2009 SRC TECHCON.

Shrinivas Venkatraman received the B.S. degree in electrical engineering from University of Pune, Pune, India, in 2004, and the M.S. degree in electrical and computer engineering from University of Minnesota, Minneapolis, in 2007.

Since 2007, he has been pursuing his career as a Design Engineer with Intel Corporation, Folsom, CA, developing Intel's next generation Microprocessors.

Mr. Venkatraman was a recipient of an award for the 2009 DAC/ISSCC Student Design Contest.

Paulo F. Butzen (SM'10) received the B.S. degree in computer engineering and the M.S. degree in computer science from Instituto de Informática-UFRGS, Federal University of Rio Grande do Sul, Porte Alegre, Brazil, in 2004 and 2007, respectively, where he is currently pursuing the Ph.D. degree in microelectronics.

He was with the VLSI Research Group, University of Minnesota, Minneapolis, in 2006. In 2007, he worked as a Design Engineer with Nangate Inc. His current research interests include the analysis, optimization and design of low-power and reliable circuits using nanoscaled CMOS technologies.

Chris H. Kim (M'04) received the B.S. degree in electrical engineering and the M.S. degree in biomedical engineering from Seoul National University, Seoul, Korea, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN.

He spent a year at Intel Corporation, where he performed research on variation-tolerant circuits, on-die leakage sensor design and crosstalk noise analysis. He joined the Electrical and Computer Engineering Faculty, University of Minnesota, Minneapolis, in 2004, where he is currently an Associate Professor. He is an author/coauthor of over 60 journal and conference papers and has served as a technical program committee member for numerous circuit design conferences. His current research interests include digital, mixed-signal, and memory circuit design for silicon and non-silicon technologies.

Prof. Kim was the recipient of the NSF CAREER Award, Mcknight Foundation Land-Grant Professorship, 3M Non-Tenured Faculty Award, DAC/ISSCC Student Design Contest Awards, IBM Faculty Partnership Awards, IEEE Circuits and Systems Society Outstanding Young Author Award, ISLPED Low Power Design Contest Awards, Intel Ph.D. Fellowship, and Magoon's Award for Excellence in Teaching.