# Compact Degradation Sensors for Monitoring NBTI and Oxide Degradation

Prashant Singh, Eric Karl, *Member, IEEE*, David Blaauw, and Dennis Sylvester, *Fellow, IEEE*

***Abstract*—We designed two compact *in situ* NBTI and oxide degradation sensors with digital outputs in 130 nm CMOS. The 308 $\mu m^2$ NBTI sensor and the 150 $\mu m^2$ oxide degradation sensor provide digital frequency outputs and are compatible with a cell-based design methodology without requiring analog supplies. The sensors enable high-volume data collection and monitoring of degradation mechanisms to guide dynamic control schemes and warn of impending device failure. Large scale data-collection permits improved modeling and the potential for insight into the underlying reliability mechanisms. The oxide degradation sensor monitors the change in gate leakage under stress conditions and is the first proposed of its kind. The NBTI sensor is 110× smaller than previous work and is designed to compensate for temperature variations during measurement. A maximum error of 2.2% is observed for the NBTI sensor under process, voltage, and temperature variations. It provides $\Delta V_{\rm th}$ measurement with $3\sigma$ accuracy of 1.23 mV from 40°C–110°C.**

***Index Terms*—CMOS, negative bias temperature instability (NBTI), oxide breakdown, reliability, sensors.**

## I. Introduction

SEMICONDUCTOR reliability is a growing issue as device-critical dimensions shrink and transistor integration continues to roughly double every 24 months. Aggressive oxide thickness scaling has led to large vertical electric fields in MOSFET devices in which oxide breakdown is a critical issue. These high fields also lead to significant threshold voltage shift over time due to the negative bias temperature instability (NBTI) effect, creating additional uncertainty in device behavior. In the presence of these degradation mechanisms, it is increasingly difficult to ensure the reliability of ICs over their lifetimes. Since the sensitivity of device lifetime to operating conditions has increased, dynamic control schemes that modulate the voltage, sleep state, and workload of processing elements and circuitry in large systems [1], [2] have been proposed. Dynamic control further complicates *a priori* reliability qualification and makes a case for on-chip structures to be used for real-time estimation of device and circuit degradation [3], [4]. Since the degradation is a statistical process, hundreds or even thousands of sensors are required to estimate bounds on overall chip performance degradation [5]. Hence it is essential that the sensors are small with low power consumption.

This paper introduces two new compact structures to quantify the change in performance of devices undergoing NBTI and defect-induced oxide breakdown [6]. The small size of the sensors makes them amenable to use in a standard cell design with minimal area and power overhead. Compact sensors can be implemented in large numbers to collect high-volume data on device degradation. For instance, based on results from the test chip in this work we observe the effect of initial threshold voltage on the $V_{\rm th}$ shift due to NBTI and the correlation between amount of $V_{\rm th}$ shift and the amount of recovery, among other effects. The proposed oxide degradation sensor likewise yields statistical data concerning initial and post-degradation gate leakage that can be useful in device modeling and circuit design.

The remainder of this paper is organized as follows. Section II describes the NBTI degradation mechanism and previous work aimed at NBTI characterization. Section III provides the same background for gate oxide breakdown. Section IV explains the design and circuit operation of the NBTI sensor and discusses the experimental results from the test chip. Section V contains similar details for the proposed oxide breakdown sensor. Finally, Section VI summarizes the work.

## II. Negative Bias Temperature Instability

### A. Degradation Mechanism

NBTI results in an increased absolute threshold voltage of p-channel MOSFETs, and hence a degradation in drain current and performance. Although NBTI is not a new phenomenon, it has recently become a major reliability issue due to high gate electric fields resulting from scaling, high operating temperatures due to large power consumption on-chip, and the addition of nitrogen to thermally grown $SiO_2$ (since hydrogen diffusion is enhanced in nitride-oxides) [7]. Most research on NBTI attributes the threshold shift to two mechanisms. The first mechanism involves interface traps and oxide charge formation due to negative gate bias at elevated temperatures [see Fig. 1(a)]. The mechanism involves breaking of Si-H bonds at the $Si/SiO_2$ interface by a combination of electric field, temperature, and holes. It results in dangling bonds or interface traps at that interface and positive oxide charge that may be due to $H^+$ [8]. The threshold voltage shift due to this mechanism is permanent and cannot be recovered upon removal of stress. The second mechanism involves hole-trapping due to electric field in the

Manuscript received October 13, 2010; revised February 27, 2011; accepted June 10, 2011. Date of publication August 12, 2011; date of current version July 05, 2012. This work was supported in part by GSRC. Mosis provided the fabrication support.

P. Singh was with the University of Michigan, Ann Arbor, MI 48105 USA. He is now with Nvidia, Santa Clara, CA 80305 USA (e-mail: prsingh@umich.edu).

E. Karl is with Intel, Hillsboro, OR 97124 USA (e-mail: drerickarl@gmail.com).

D. Sylvester and D. Blaauw are with the University of Michigan, Ann Arbor, MI 48105 USA (e-mail: blaauw@umich.edu; dennis@eecs.umich.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2011.2161784

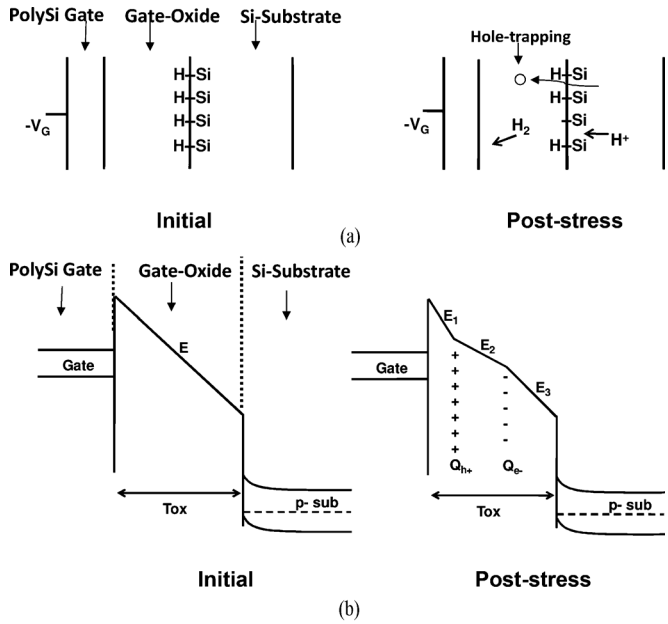


Fig. 1. (a) NBTI degradation mechanism. Interface traps are created by breaking Si-H bonds due to negative gate voltage stress. Positively charged interface traps contribute to threshold voltage shift. Hole-trapping also contributes to $V_{\rm th}$ shift. (b) Oxide degradation mechanism. Electrons flowing through the oxide generate defects which increase the local electric field in the oxide, causing more electrons to tunnel through that region and generating new defects. This positive feedback eventually causes the defects to reach a critical density, leading to destructive breakdown.

gate-oxide. Upon removal of the negative stress there is an immediate partial recovery from the threshold shift that occurred during stress. It is due to hole-detrapping in the gate-oxide [8].

NBTI poses a serious threat to chip reliability since significant $\Delta V_{\rm th}$ can lead to marginal circuit operation (e.g., timing failures in digital logic or bias point drift in analog circuitry) [9], [10]. Additionally, NBTI degradation in pMOS devices has been shown to cause a reduction in the static noise margin of SRAM cells, leading to read-induced cell stability issues [11].

NBTI recovery adds to the complexity of measuring NBTI effects. If the measurement interrupts the stress state of the device, the measurement time needs to be sufficiently small (on the order of ms) to avoid masking the actual $V_{\rm th}$ shift by inadvertent recovery [12]. This poses the biggest challenge to any experiment seeking to characterize NBTI. The recovery effect also makes it difficult to assess the lifetime of a device undergoing NBTI since the behavior of a device under stress changes once it has been subjected to recovery.

### B. Previous Measurement Techniques

The aim of most previous work in NBTI measurement has been in the direction of characterizing NBTI. In the past, researchers have used invasive probing methods that require direct access to the device-under-test (DUT) to monitor currents. One large class of work [13]–[19] employs a direct current probing approach. Ring oscillator based structures have been proposed in [4], [20], and [21]. All these structures consist of a pair of ring oscillators, one of which experiences accelerated stress. The structure proposed in [4] and [20] measures the beat frequency (i.e., difference of the two oscillator frequencies) which is attributed to the $V_{\rm th}$ shift due to NBTI. These structures produce a digital output which makes it easier to collect and process it. The structure proposed in [21] requires a controllable external analog bias to map the change in beat frequency to $V_{\rm th}$ shift. In [3] this analog bias is generated on-chip using a delay-locked loop and the ring oscillators is replaced with a voltage controlled delay line. The analog output makes it harder to collect the data from these two approaches [3], [21]. All of these four proposed structures [3], [4], [20], [21] require a large number of delay stages to get high sensitivity to $V_{\rm th}$ change and hence are costly in terms of area and not ideal for use in large numbers like hundreds or thousands as on-chip sensors.

## III. Gate Oxide Breakdown

### A. Degradation Mechanism

When a voltage is applied across the gate oxide of a MOSFET, current flows through it if the gate voltage is high enough or if the oxide is thin enough. Fowler-Nordheim tunneling controls the current for thick oxides [22], [23], while for thin oxides ($t_{\rm ox} \leq 3$ nm) at voltages less than roughly 3 V (corresponding to the barrier height between n-type silicon and $SiO_2$) the quantum-mechanical tunneling causes the current flow. Electrons flowing through the oxide generate defects at a rate depending on their energy [see Fig. 1(b)]. These defects increase the local electric field in the oxide, causing more electrons to tunnel through that region and generating new defects. This positive feedback eventually causes the defects to reach a critical density, leading to destructive breakdown [24] where the oxide ceases to behave as an insulator.

### B. Previous Measurement Techniques

As in the case of NBTI, all previous oxide breakdown measurement techniques have been invasive, requiring direct access to DUTs and peripheral circuitry. Uraoka [25] evaluates gate oxide reliability using a luminescence method. The set-up requires an optical microscope, photon counting camera, and image processor. [26] and [27] propose statistical methods to monitor the yield of gate oxide layers in a manufacturing production line. This technique can be useful in statistically binning the oxide reliability of ICs at manufacturing time but cannot dynamically monitor chip reliability throughout its lifetime.

Gate oxide reliability for high voltage analog power transistors is addressed in [28]. A power amplifier (PA) is designed with an oxide reliability monitor for the output stage. The monitoring circuit uses the elements already present in the PA, such as resistors and a pre-driver stage, to measure the conductance of output stage transistors. Extending this approach to digital circuits would result in large silicon area and power overheads due to analog components. Also, the output of the monitor is an analog voltage, increasing testing costs. Reference [29] proposes an array-based test structure to statistically characterize gate oxide breakdown. The area for 1024 oxide-arrays along with analog and digital blocks is approximately 0.7 $\rm mm^2$. Such structures can be used to dynamically monitor oxide degradation if its area overhead can be reduced.

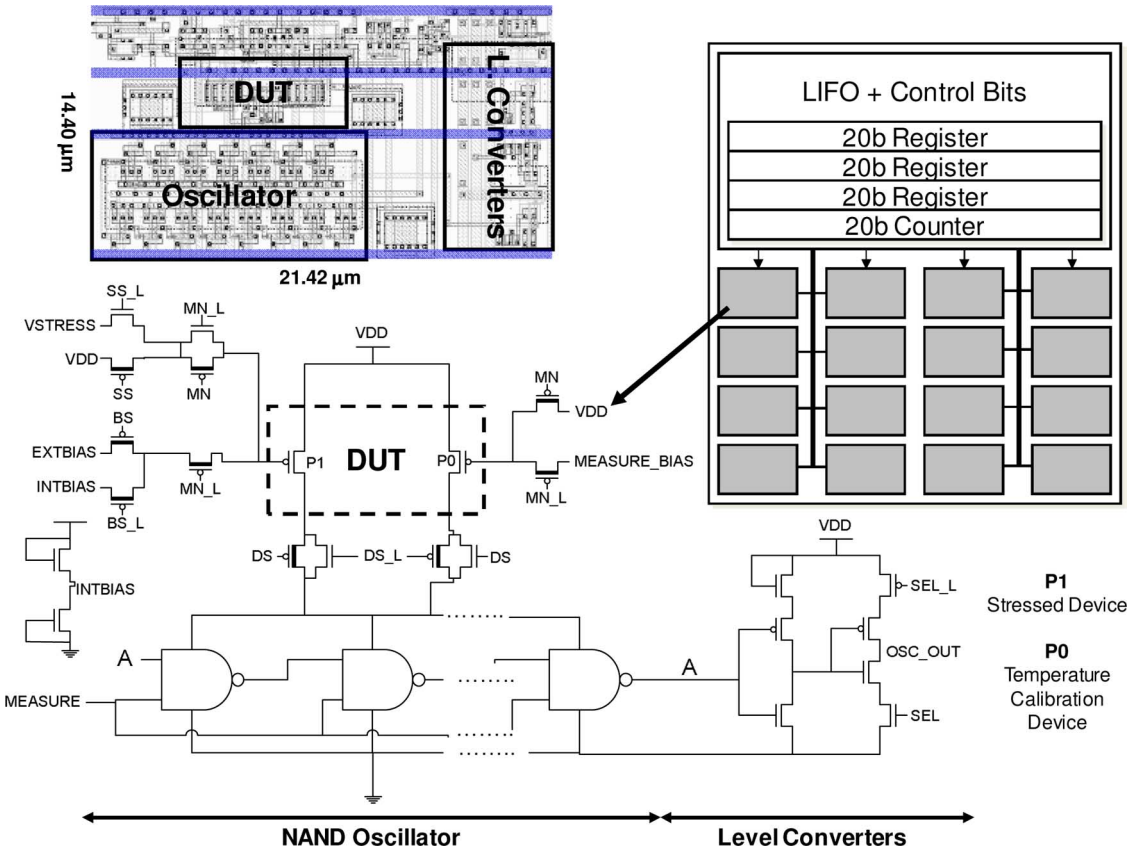


Fig. 2. Circuit schematic and layout for the proposed NBTI sensor and the block diagram for a bank of sensors on the test chip. The sensor operates in i) Measurement mode or ii) Stress/Recovery mode. In i) P1/P0 are biased by EXTBIAS or INTBIAS ($\sim 1\mathrm{V}$), while in ii) P1 is biased with VSTRESS ($\leq 0$) or VDD (recovery mode) and P0 is biased with VDD.

## IV. NBTI Sensor

### A. Circuit Design Principles

Fig. 2 shows the full schematic of the proposed NBTI sensor, its layout, and the placement scheme for sensors on the test chip. The NBTI measurement technique relies on a pMOS device (P1) to starve the current supplied to a 15-stage NAND gate ring oscillator. There are three modes of operation for this sensor: stress mode, measurement mode, and recovery mode. In the stress mode, P1 is stressed with negative bias by grounding the input. The change in oscillation frequency during the lifetime of the sensor quantifies the change in $V_{\mathrm{th}}$ of P1. In the measurement mode, P1 is biased in subthreshold to exponentially sensitize the oscillation frequency to $\Delta V_{\mathrm{th}}$. Experimental results show that biasing P1 in subthreshold leads to a 53% change in oscillator frequency $(f_{\mathrm{osc}})$ for 10% $\Delta V_{\mathrm{th}}$. During recovery mode, the gate of P1 is tied to VDD to allow NBTI recovery.

Subthreshold current is highly sensitive to temperature; therefore, we implemented a control pMOS header P0 to correct for temperature variation. Any change in the P0-starved oscillation frequency gives a measure of this variation. A mathematical model, explained in Section IV-B, maps $\Delta V_{\mathrm{th}}$ and temperature to oscillation frequency. The estimated temperature and the P1-starved oscillation frequency are used to quantify the $\Delta V_{\mathrm{th}}$ of P1 after stress.

The additional circuitry consists of an internal bias generator (for measurement mode), four multiplexers (to switch between different modes of operation), and a level converter (at the oscillator output). Multiplexers selectively put: 1) P1 in stress or recovery mode (which allows the application of AC stress on P1), while P0 is always in the unstressed mode; 2) enable switching from stress/recovery mode to measurement mode; 3) select between the internal bias generator and external bias (from pads); and 4) select either P0 or P1 to starve the oscillator. Since the oscillator is strongly current starved, the oscillation amplitude is small. Thus a level converter is used to restore the amplitude of oscillations. The sensors are arranged in an array and form a bank. The bank also includes one 20-bit counter and three subsequent serially connected 20-bit storage units. The counter and three storage units together allow four quick measurements when a stress cycle is interrupted to conduct measurements, and allow quantifying the fast recovery process. In total, a die contains 96 NBTI sensors.

### B. Oscillator Modeling

The oscillation frequency of the NBTI sensor oscillator can be approximately formulated as

$$\frac{1}{f_{\mathrm{osc}}} = \frac{\mathrm{KCV}_{\mathrm{amp}}}{I_{\mathrm{OUT}}} \tag{1}$$

where $K$ is a fitting constant, $C$ is the capacitance at the output node of each NAND stage of the oscillator, $V_{\mathrm{amp}}$ is the amplitude of oscillations, and $I_{\mathrm{DUT}}$ is the average current through the DUT. Simulations show that the peak charging current, $I_{\mathrm{charge}}$,

for a single NAND stage of the ring oscillator is roughly equal to $I_{\rm DUT}$. This helps to arrive at an expression for $V_{\rm amp}$. Since P0 is strongly starved so that the ring oscillator operates in subthreshold, $I_{\rm DUT}$ and $I_{\rm charge}$ are governed by the subthreshold current equation

$$I_{\rm sub} = AT^2 e^{V_{CS}-V_{\rm TH}-\gamma V_{BS}+\eta V_{DS}/nT}\left(1-e^{-V_{DS}/VT}\right). \quad (2)$$

Here $A$ is a constant that is dependent on device parameters $W$, $L$, $\mu$, and $C_{\rm ox}$.

Using (2) for the DUT and (3), shown at the the bottom of the page, where $V_{\rm bias}$ is the bias voltage at the gate of the DUT. Since $V_{BS}$ is 0 for the DUT (Body and Source are at Vdd), the $\gamma V_{BS}$ term can be dropped. Also, with $V_{DS} \gg V_T$ ($V_{DS}$ varies from 0.7 to .9 V for DUT, $V_T \sim 35$ mV at 100 °C) the last term $\left(1-e^{-V_{DS}/VT}\right)$ can be dropped. Hence

$$I_{\rm DUT} = A_{\rm DUT} T^2 e^{V_{\rm DD}-V_{\rm bias}-V_{\rm thDUT}+\eta(V_{\rm DD}-V_{\rm amp})/nT}.$$

Similarly the equation for the current sinking into oscillator can be framed. Again for the oscillator-transistors $V_{BS}$ is 0 so the $\gamma V_{BS}$ term can be dropped. $V_{DS} \gg V_T$ ($V_{DS}$ varies from 0.5 to 0.3 V for ring oscillator transistors, $V_T \sim 35$ mV at 100 °C) the last term $\left(1-e^{-V_{DS}/VT}\right)$ can be dropped. After making these simplifications following equation can be written:

$$I_{\rm charge} = A_{\rm charge} T^2 e^{V_{\rm amp}-V_{\rm thROSC}+\eta V_{\rm amp}/nT} \quad (4)$$

where $V_{\rm thROSC}$ refers to the effective $V_{\rm th}$ of the devices in the ring oscillator.

Based on the previous argument

$$I_{\rm DUT} = I_{\rm charge}. \quad (5)$$

Using (3)–(5), the expression for $V_{\rm amp}$ becomes

$$V_{\rm amp}(V_{\rm thDUT}, T) = A_1 T + A_2 V_{\rm thDUT} + A_3 \quad (6)$$

where

$$A_1 = \frac{n}{1+2\eta}\ln\left(\frac{A_{\rm DUT}}{A_{\rm charge}}\right)$$
$$A_2 = -\frac{1}{1+2\eta}$$
$$A_3 = \frac{V_{\rm DD}(1+\eta) - V_{\rm bias} + V_{\rm thROSC}}{1+2\eta}.$$

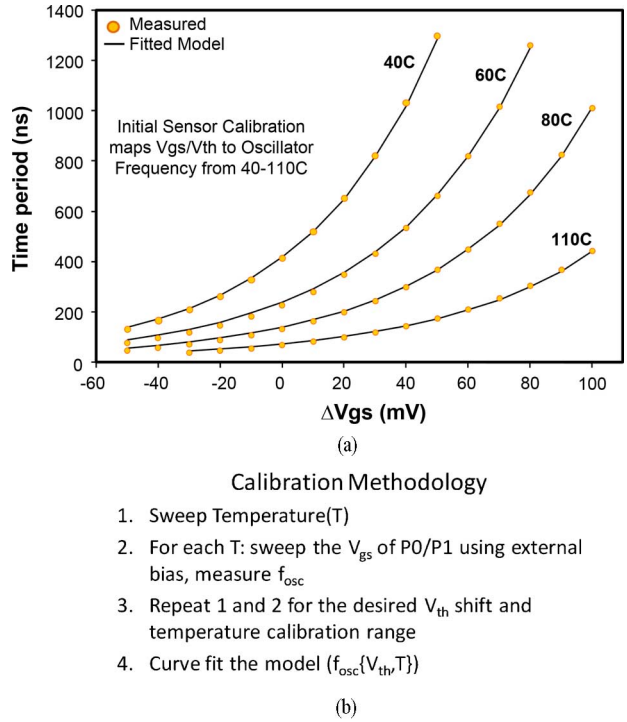


Fig. 3. (a) Discrete experimentally measured frequency and frequency given by calibrated model versus emulated $\Delta V_{\rm th}$ at several temperatures. The model fits the experimental data to minimize the actual $\Delta V_{\rm th}$ and $\Delta V_{\rm th-predicted}$ by sensor. (b) Overview of the calibration methodology.

Next we model the threshold voltage as a function of temperature. $V_{\rm thROSC}$ can be modeled as

$$V_{\rm thROSC}(T) = V_{\rm tho} - T_{\rm coff}\log\left[\frac{T}{T_L}\right] \quad (7)$$

where $V_{\rm tho}$ is the threshold voltage at $T = T_i$. The value for $V_{\rm tho}$ and the corresponding $T_i$, are taken from SPICE device models. Equation (1), (4), (6), and (7) form a system of equations to give (8), as shown at the bottom of the page. Here $V_{\rm tho}$ is taken from SPICE models, $V_{\rm thDUT}$ is the threshold voltage of the DUT, $T$ is the temperature, and $A, \eta, n, A_1, A_2, A_3$ are fitting parameters.

### C. Calibration and Measurement Strategy

The mathematical model developed in the last section must be calibrated, i.e., the fitting parameters need to be calculated. The

$$I_{\rm DUT} = A_{\rm DUT} T^2 e^{V_{\rm DD}-V_{\rm bias}-V_{\rm thDUT}-\gamma V_{BS}+\eta(V_{\rm DD}-V_{\rm amp})/nT}\left(1-e^{V_{DS}/VT}\right) \quad (3)$$

$$f_{\rm OSC} = AT^2 e^{A_1 T+A_2 V_{\rm thDUT}+A_3(1+\eta)-V_{\rm tho}/nT}\frac{1}{A_1 T + A_2 V_{\rm thDUT} + A_3} = g(V_{\rm thDUT}, \{A, \eta, n, A_1, A_2, A_3\}). \quad (8)$$

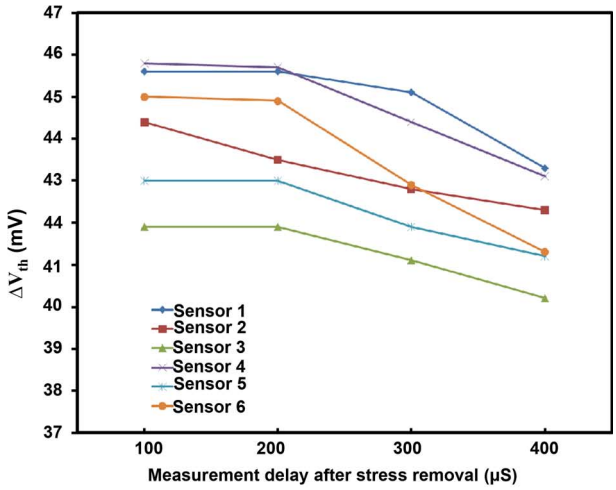


Fig. 4. (a) Effect of measurement time on $V_{\rm th}$ recovery. (b) Comparison of sensor output with internal and external bias. Internal bias incurs a maximum error of 2.2% over the external bias for the shown sample.

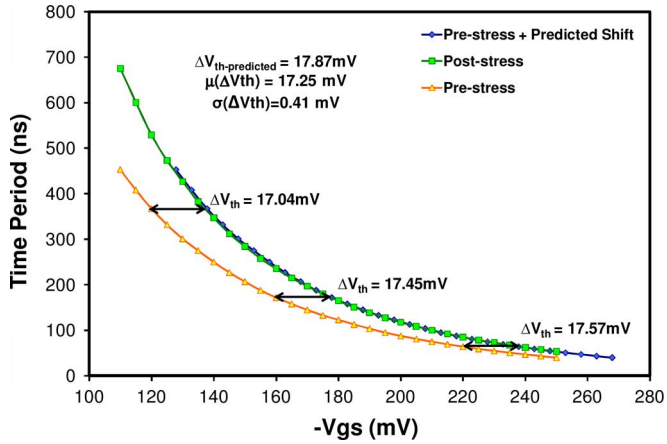


Fig. 5. Verification of NBTI measurements: $V_{\rm gs}$ for P1 is swept before stress and after stress under constant conditions of $V_{\rm dd}$ and $T$; the horizontal shift in the two profiles is equal to $\Delta V_{\rm th-sweep}$. The pre-stress profile shifted horizontally by $\Delta V_{\rm th-predicted}$ overlaps very well with the post-stress profile.

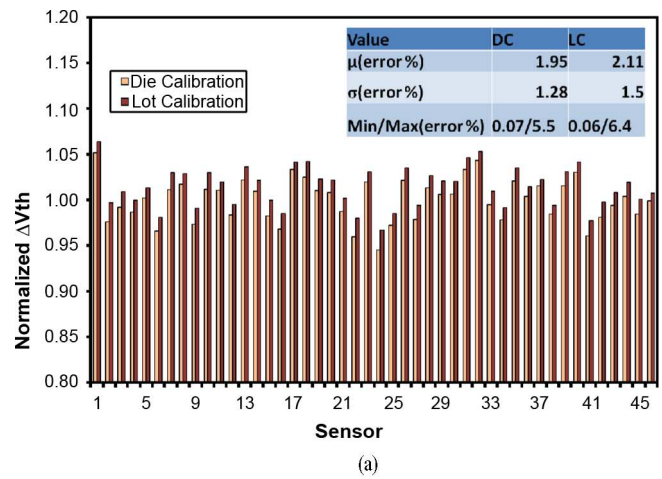


Fig. 6. (a) Computation of $\Delta V_{\rm th}$ using individual calibration (IC), die calibration (DC), and lot calibration (LC) for 46 sensors on a typical die. The maximum error incurred by lot calibration $\sim$ 6%, which makes it quite reliable to use and simplifies the calibration procedure drastically. (b) Temperature sensitivity of different calibration methods. Individual Calibration varies the least across the temperature range, followed by Die Calibration and then Lot Calibration.

calibration steps are shown in the table of Fig. 3(b). The temperature and gate bias of the DUT are swept, and frequency is measured. The effect of sweeping gate bias is equivalent to varying threshold voltage. This data is used to curve-fit the model given by (8). Results of the curve fitting at different temperatures are also shown in Fig. 3(a). The values of the fitting parameters will differ across DUTs due to process variation. Using a particular fixed set of parameters for all DUTs leads to error in the computation of temperature and $\Delta V_{\rm th}$. This error varies with the calibration methodology employed. This work investigates three different methodologies and examines the resulting error in each case: 1) individual; 2) die; and 3) lot calibration. Using individual calibration, the fitting parameters are calculated separately for each DUT. This gives the minimum error in temperature and $\Delta V_{\rm th}$ but at the cost of greatly increased test time and post-test computational requirements. Die calibration involves calibrating the model for only one DUT on each die and then using these derived parameters to model all other DUTs on the chip. With small intra-die variations, this method delivers very reasonable accuracy with a significant reduction in test/calibration time. With lot calibration, just one DUT is examined for the entire wafer lot and fitting parameters from that device are used

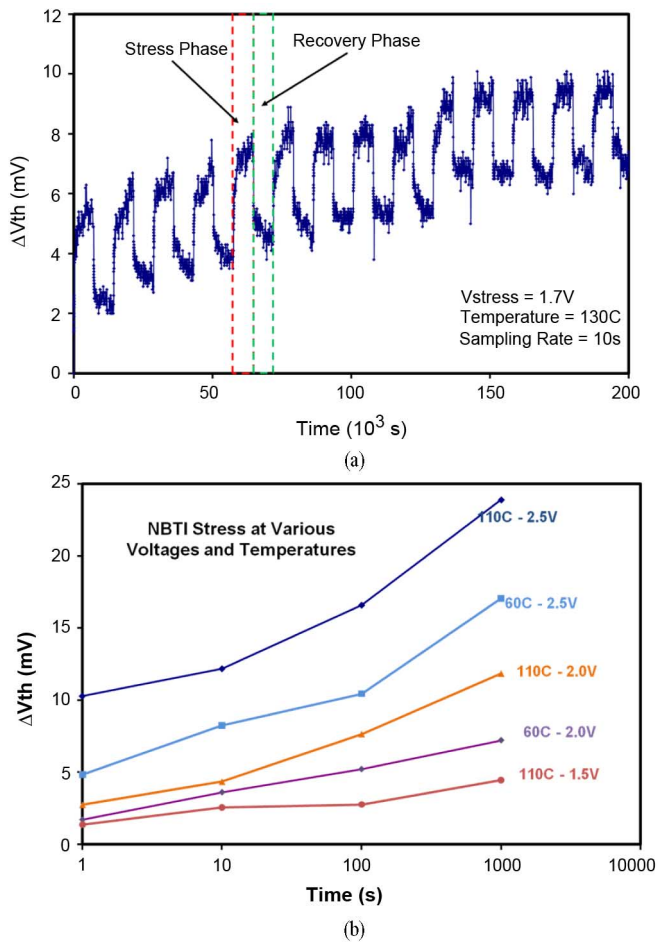


Fig. 7. (a) NBTI measurement for 15 cycles of periodic stress and recovery mode. (b) NBTI measured from the sensor under various stress conditions.

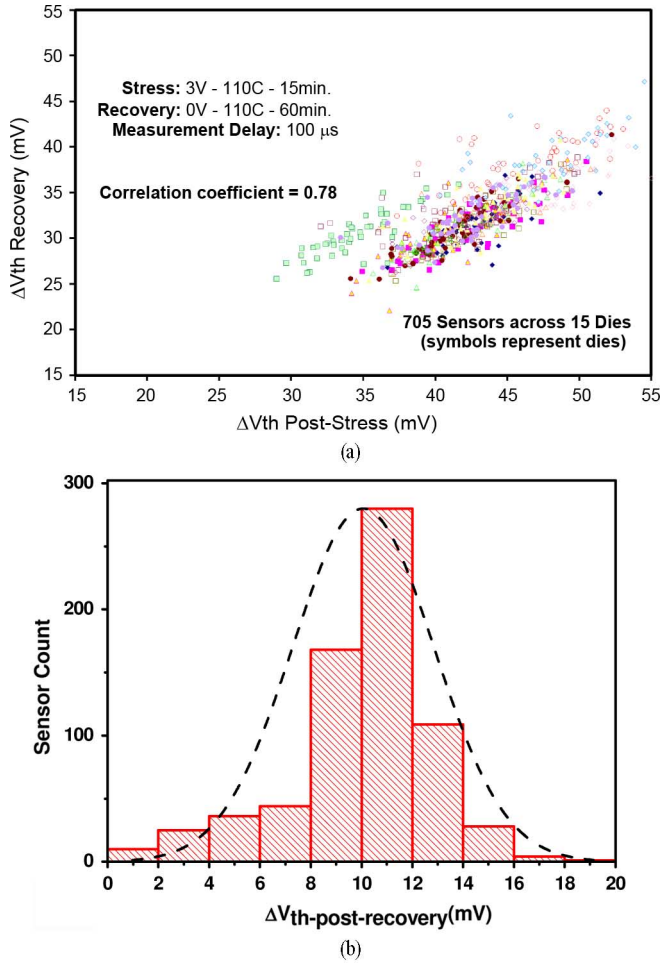


Fig. 8. (a) Correlation between amount of $\Delta V_{\rm th}$ post-stress and the amount of recovery. (b) Distribution of $\Delta V_{\rm th}$ post recovery.

for other DUTs within that lot. Experiments were performed to quantify the incurred error when different calibration methodologies are adopted; the results are discussed in the following section.

### D. Experimental Results

The test chip, fabricated in 1.2 V/3.3 V 130-nm CMOS technology, contains 96 NBTI sensors arranged in six banks. Each bank contains 16 NBTI sensors, which are individually addressable for measurement using a 20-bit counter. The threshold voltage in this technology has a nominal value of 355/-325 mV for nMOS and pMOS devices, respectively.

The area of each NBTI sensor is 308 $\mu$m$^2$ with a stress mode power consumption of 4.5 nW and a measurement power of 500 nW. A measurement time of 100 $\mu$s is achieved for results presented in this work. The three serially connected 20-bit storage registers were used to perform four quick $V_{\rm th}$ measurements at the interval of 100 $\mu$s, to see the effect of the measurement time on amount of recovery induced by measurements. The results shown in Fig. 4(a) indicate that the measurement period of 100 $\mu$s induces minimal $V_{\rm th}$ recovery. Further, the measurement time can be vastly improved by using a high-speed clock for the counter and using the sensor output as an enable signal for the counter. This high speed clock can be as slow as 500 MHz to give a high counter count. This clock can be easily tapped off the system clock with miniscule area and power overhead over the system's area and power consumption. An internal bias generator is included in the design that incurs a maximum of 2.2% error in $\Delta V_{\rm th}$ measurement relative to the external bias [see Fig. 4(b)]. The external bias is used for the results presented here. Fig. 5 shows results from an experiment to verify that the proposed method of NBTI $\Delta V_{\rm th}$ estimation measures the actual shift in $V_{\rm th}$ of the device P1.

In this experiment, the gate voltage of the DUT is swept, showing time-period versus $V_{\rm gs}$ (Curve A). The DUT is then stressed for some time and then allowed to recover. After the recovery saturates, the gate voltage of the DUT is again swept to obtain Curve B. The horizontal shift between the two curves gives the residual $\Delta V_{\rm th-sweep}$. Then the sensor is used to find $\Delta V_{\rm th-predicted}$ based on oscillation frequency. The sensor predicts $\Delta V_{\rm th}$ accurately with a $3\sigma$ error of 1.23 mV. Thirteen different points on Curves A and B are used to compute the mean and standard deviation for $\Delta V_{\rm th-sweep}$.

We also measured the error introduced by using die or lot calibration methodologies. Here the error is computed and normalized with respect to individual calibration. As shown in Fig. 6(a), $\mu(\Delta V_{\rm th-error}) = 1.95\%$ and $\sigma(\Delta V_{\rm th-error}) = 1.28\%$ for die calibration and $\mu(\Delta V_{\rm th-error}) = 2.11\%$ and

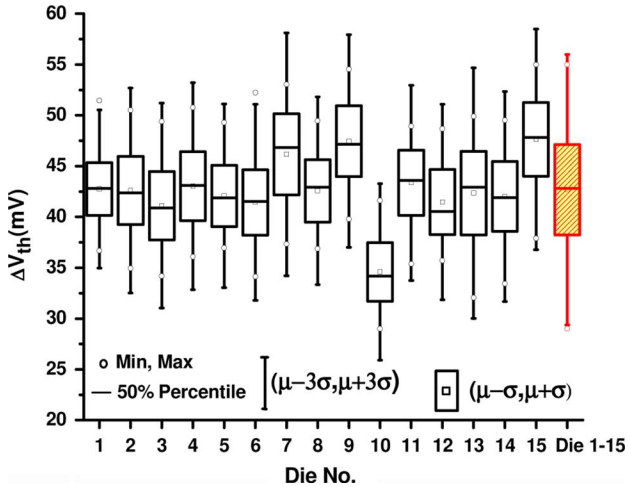


Fig. 9. Intra-die and global variation for 15 dies.

TABLE I
MEAN AND STANDARD DEVIATION OF $\Delta V_{\rm th}$ FOR 15 DIES

| Die no. | μ(Δ Vth) (mV) | σ(Δ Vth) (mV) |
|---|---|---|
| 1 | 42.74787234 | 2.59707941 |
| 2 | 42.60326087 | 3.357749282 |
| 3 | 41.11041667 | 3.364005944 |
| 4 | 43.02913043 | 3.393326671 |
| 5 | 42.07382979 | 3.010861175 |
| 6 | 41.4275 | 3.215041444 |
| 7 | 46.15893617 | 3.987418929 |
| 8 | 42.566875 | 3.081232246 |
| 9 | 47.46043478 | 3.487033095 |
| 10 | 34.59340426 | 2.895410369 |
| 11 | 43.3512766 | 3.199630833 |
| 12 | 41.46395833 | 3.204979994 |
| 13 | 42.34354167 | 4.112595851 |
| 14 | 42.00958333 | 3.445013021 |
| 15 | 47.63090909 | 3.620797046 |
| Die 1-15 | 42.67 | 4.44 |

$\sigma(V_{\rm th-error}) = 1.5\%$ for lot calibration. The confidence interval analysis shows that the margin of error in estimation of mean for die calibration is 0.31% and for lot calibration is 0.36% with confidence level of 90%.

These small errors imply that die and lot calibration provide a good tradeoff between accuracy and calibration/testing effort. This has enormous implications on the usability of these sensors in actual chips. For a particular process-technology the sensor can be extensively calibrated and that calibration can be used for the sensors deployed on chips in that process-technology. In that view the sensors would not need an external bias (primarily required for calibration process) which will reduce the routing overhead of this signal all over the chip. This study was extended to test for temperature correction [see Fig. 6(b)] and measurements were taken for a wide range of temperatures (10 °C to 110 °C). The estimates derived from the individual calibration methodology vary the least $(\sigma/\mu = 1.1\%)$ across the range, followed by die calibration $(\sigma/\mu = 2.1\%)$ and then lot calibration $(\sigma/\mu = 3.1\%)$. Fig. 7(a) shows the characteristic saw-tooth curve generated by alternately stressing and then removing the

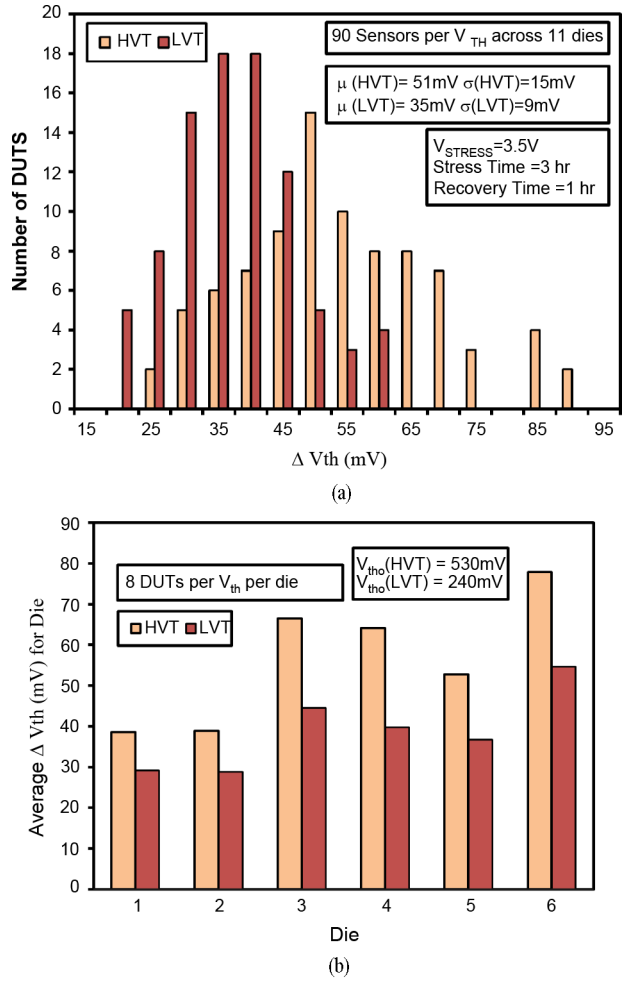


Fig. 10. (a) Distribution of NBTI-induced Vth shift across DUTs from 11 dies. HVT devices show higher degradation. (b) The mean $V_{\rm th}$ shift of HVT devices is larger across dies. The respective $V_{\rm th}$ of devices are 530 mV (HVT) and 240 mV (LVT).

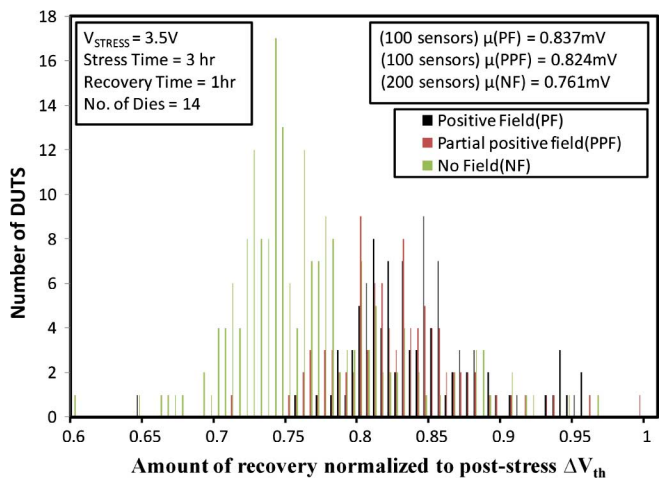


Fig. 11. Distribution of amount of recovery for DUTs recovering under different gate source/drain electric field conditions. Application of a positive field increases the amount of recovery in a given time.

stress with $V_{\rm stress} = 1.7$ V and $T = 130$ °C. The jitter in oscillations of the sensor due to power supply and bias voltage noise

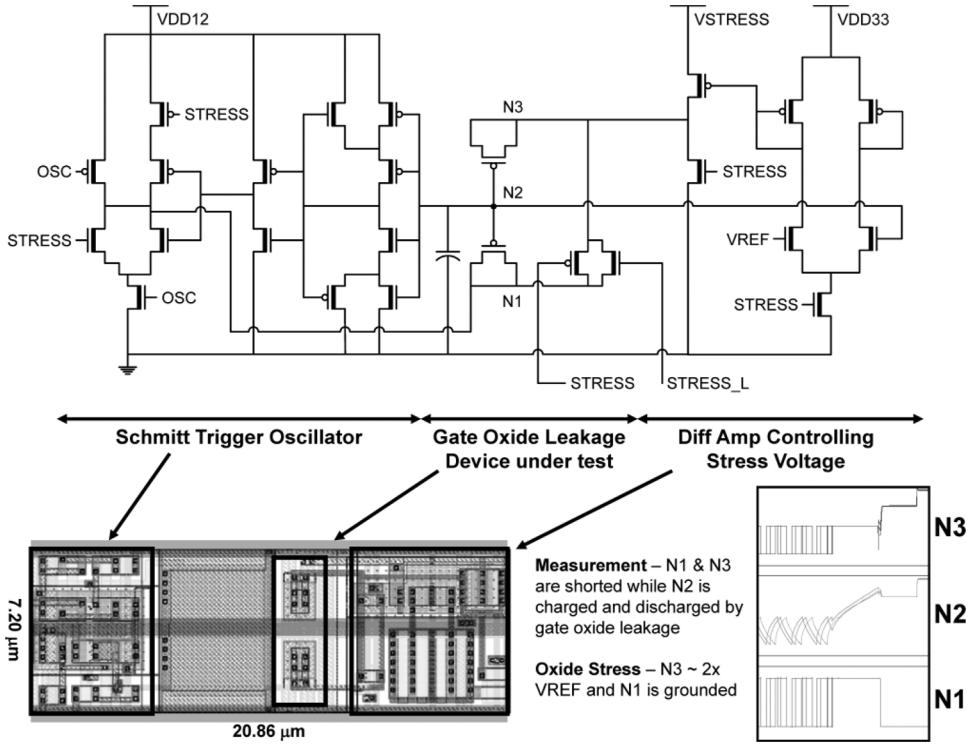


Fig. 12. Circuit schematic and layout for the proposed oxide degradation sensor. Gate leakage through pMOS devices between N1/N3 and N2 controls the sensor oscillation frequency. High voltage is applied at N3 and ground to N1 during the stress mode of operation.

results in noise in the $\Delta V_{\rm th}$ estimate. Fig. 7(b) also shows the NBTI degradation for different stress conditions.

The statistical nature of NBTI degradation is one aspect that has been largely overlooked in the literature. To investigate the probabilistic nature of NBTI, we collected stress/recovery data for 705 sensors across 15 dies with $V_{\rm stress} = 3$ V and $T = 120\,^{\circ}$C [see Fig. 8(a)]. A measurement time of 100 $\mu$s is used. We observed a strong positive correlation (correlation coefficient = 0.78) between the amount of $\Delta V_{\rm th}$ during stress and the amount of recovery. Fig. 8(b) shows the distribution of $\Delta V_{\rm th}$ post recovery. This shows that the devices suffering more NBTI degradation also recover more significantly if given sufficient recovery time. This finding can be used to dynamically mitigate the severity of NBTI by allocating sufficient recovery time for the blocks suffering higher threshold shifts.

Fig. 9 shows intra-die distribution of $\Delta V_{\rm th}$ across 15 dies after NBTI stress and the global distribution. The intra-die variation of $\Delta V_{\rm th}$ across the dies is fairly consistent. The mean value of four dies (die 7, 9, 10, 15) is off from the cluster of mean $\Delta V_{\rm th}$ cluster of other dies. The intra-die variation is much more pronounced as compared to inter-die variation. Due to this the global distribution looks very much similar to the intra-die distribution. Table I shows the mean and standard deviation of $\Delta V_{\rm th}$ for each die and the global distribution. The statistical data shows that the impact of NBTI on circuit performance is more complex than previously considered. These findings also make the traditional reliability solution, like delay margining, more pessimistic and corroborate the case for dynamic reliability monitoring using NBTI sensors.

The test chip also allowed the investigation of DUTs with two different (high and low) threshold voltages to examine differences in their NBTI characteristics. Fig. 10(a) and (b) show that high-$V_{\rm th}$ DUTs exhibit larger NBTI degradation overall. A possible explanation could be based on the NBTI model proposed by Tsetseris *et al.* [30]. According to this model the following chemical reaction occurs at the $\rm Si - SiO_2$ interface under stress conditions:

$$\rm Si_3 \equiv SiH + H^+ \rightarrow Si_3 \equiv Si \bullet + H_2.$$

Here $\rm Si_3 \equiv SiH$ is a hydrogen-terminated interface trap and $\rm Si_3 \equiv Si\bullet$ an interface trap with the dot representing a dangling bond. The hydrogen is believed to originate from the phosphorus-hydrogen bonds in the n-type Si substrate. The P-H bonds dissociate and the hydrogen attracts a hole as it moves to the $\rm SiO_2/Si$ interface, becoming $\rm H+$, then reacts with the H from the SiH bond to form $\rm H_2$. The net result is a positively charged Si dangling bond (or trapping center) that contributes to the $V_{\rm th}$ shift. Since the phosphorus doping is larger for high $V_{\rm th}$ devices, there is a higher concentration of $\rm H+$ available for the reaction, resulting in more traps and a consequently greater $V_{\rm th}$ shift. Another possible explanation could be that due to higher doping there are more Si atoms in the lattice which are not at the Si lattice centers. These dangling Si bonds are passivated by hydrogen. Since there are more number of SiH bonds which could be broken during NBTI stress, it would lead to higher $V_{\rm th}$ shift.

We also investigated the impact of electric field on recovery by designing DUTs to recover in three different conditions of gate bias: 1) zero gate-source/drain bias; 2) positive gate-source/drain bias; and 3) positive gate-drain bias with zero gate-source bias.

Fig. 11 shows a significant difference in the mean $\Delta V_{\rm th}$ of zero and positive bias conditions, indicating that the recovery rate is enhanced in the presence of positive electric field. As explained in Section II, positive holes trapped in oxide defects also contribute to the threshold shift during negative stress. These charged species are neutralized when the negative stress is removed, which contributes to threshold voltage recovery. When a positive bias is applied at the gate, more of these trapped holes are neutralized, resulting in larger recovery.

## V. Oxide Degradation Sensor

### A. Circuit Design Principles

The proposed oxide degradation sensor consists of a pair of pMOS (M1, M2) devices connected with a small Schmitt trigger-based ring oscillator and a differential pair to drive the gate oxide with a high voltage when stressed (see Fig. 12). The frequency of oscillation is set by the gate leakage through the pMOS devices and the size of the capacitor at the node driven by gate leakage, N2. During measurement, the STRESS signal is driven low to short the drain-source-bulk nodes of both pMOS devices to the output of the oscillator and to deactivate the output of the diff amp connected to N3. The OSC signal is driven high to enable the ring oscillator to toggle, limited by the gate leakage current through the pMOS devices. To stress the oxide devices, node N2 is driven through a series resistive gate leakage divider formed by M1 and M2, while N3 is driven to a stress voltage VSTRESS and N1 is tied to ground. In this case, N2 is held to a nominal voltage which is half of the applied voltage at N3. Since an increase in leakage in one pMOS device (M1/M2) leads to an increased voltage drop in the other (M2/M1), both M1 and M2 suffer approximately same degradation and hence the voltage at node N2 remains fairly constant with time.

The scheme of two gate oxides in a stack is essential to isolate node N2 from spurious source/drain diffusion currents. During measurement, node N2 is driven solely by the gate leakage of M1/M2. In the 130-nm technology in which this circuit is implemented, subthreshold current is much larger than gate leakage, eliminating the possibility of any source/drain connected transistors at node N2. In newer technologies, it may be possible to reduce subthreshold leakage current below gate leakage current levels by using appropriate circuit techniques (e.g., super cutoff devices). In this case, the use of two stacked MOS devices is obviated, simplifying the sensor design and making it more robust. An important design constraint is the amount of current transistor M3 can source to the oxide stack once the oxides start to degrade. If the stress voltage is held constant throughout the stress period, the gate leakage will increase gradually and eventually rise dramatically upon hard breakdown. But for practical sizes of M3 (same current sourcing capability as typical core devices), the degradation might become self-limiting once the gate-current increases beyond a certain limit. Similar trend would be observed in core devices due to limited current sourcing capabilities. This allows better matching between the degradation characteristics of the sensor and core devices. Complete oxide breakdown was therefore not anticipated in the experimental results due to self-limiting nature of the gate-oxide degradation.

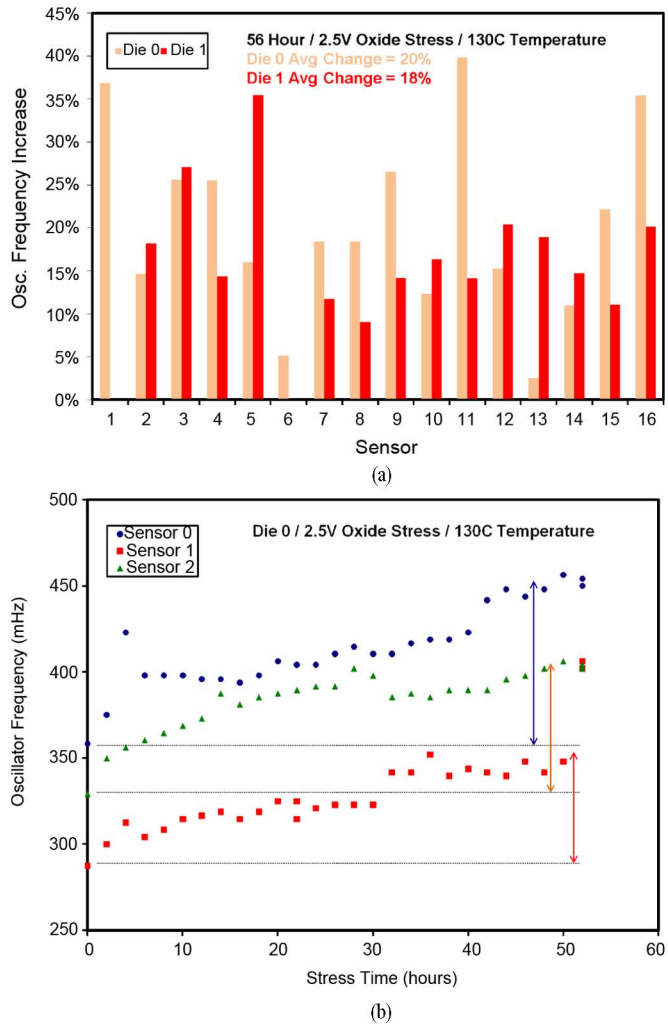


Fig. 13. (a) Oxide degradation sensor results showing relative frequency shift after 56 hour stress period. (b) Oxide degradation sensor output for three sensors at a given stress condition.

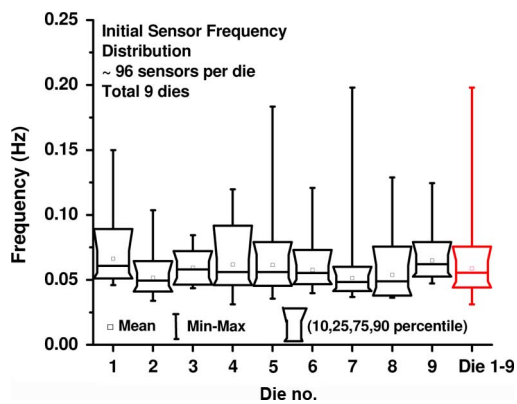


Fig. 14. Intra-die and inter-die distribution of initial gate leakage frequency.

### B. Experimental Results

The test chip contains 144 oxide degradation sensors arranged in an array structure with four banks. Each bank of

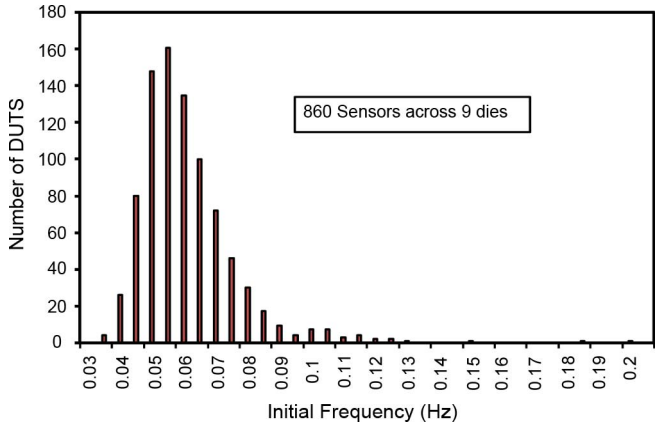


Fig. 15. Oxide degradation sensor: Cumulative initial oscillation frequency distribution across dies, providing a measure of process-induced gate leakage variability (pre-stress).

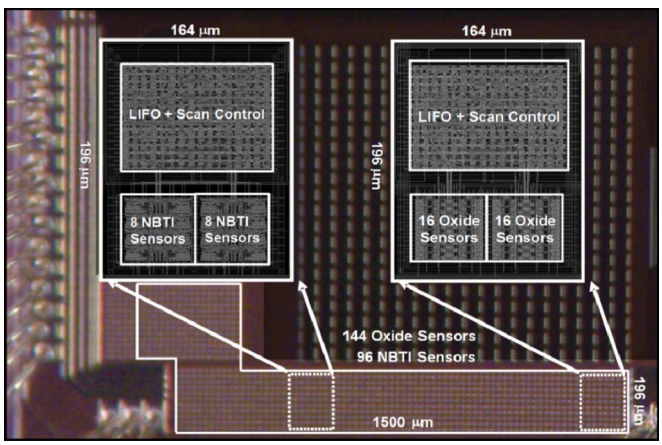


Fig. 16. Chip-microphotograph.

sensors contains between 16 and 32 oxide degradation sensors, which are individually addressable for measurement using a 20-bit counter. The nominal gate oxide thickness for standard devices is 2.2 nm.

The oxide sensor is 150 $\mu$m$^2$ and consumes 469.5 $\mu$W during stress mode and 14.03 $\mu$W during measurement mode. To accelerate oxide degradation for data collection, ambient temperatures from 130 °C–175 °C and $V_{\mathrm{STRESS}}$ voltages between 5–6 V are applied to achieve a stress voltage of 2.5–3 V across each pMOS device. Typical results from 56-h stress tests with 2.5 V across each oxide and 130 °C show a sharp initial rise in oscillator frequency followed by a steady increase, ultimately exhibiting 19% average degradation with a range of observed results between 5%–40% [see Fig. 13(a) and (b)]. The intra-die and inter-die distribution of initial sensor frequency is shown in Fig. 14 and Table II. The sensor frequency is linearly proportional to gate-leakage. The mean of the gate-leakage is fairly consistent across dies but the spread and the outliers vary from chip to chip, for example for die three Max/Min leakage $\sim 2$ whereas for die seven it is $\sim 5$. The global distribution is similar to intra-die variation. A detailed global distribution of sensor frequency is shown in Fig. 15. Since gate-leakage is exponentially dependent on gate-oxide thickness, the distribution has an extending tail towards sensors with higher gate-leakage.

TABLE II
INTER-DIE STATISTICS OF INITIAL GATE-LEAKAGE FREQUENCY

| Die no. | μ(Initial freq.) (Hz) | σ (Initial freq.)(Hz) |
|---|---|---|
| 1 | 0.066262 | 0.016835061 |
| 2 | 0.051513305 | 0.010525084 |
| 3 | 0.059267933 | 0.009794024 |
| 4 | 0.061759259 | 0.017755053 |
| 5 | 0.061458333 | 0.019183487 |
| 6 | 0.057647569 | 0.011691475 |
| 7 | 0.051394674 | 0.017638456 |
| 8 | 0.053807869 | 0.014505512 |
| 9 | 0.064907405 | 0.013761948 |
| Die 1-9 | 0.058658 | 0.015768 |

## VI. CONCLUSION

This paper presents two compact *in situ* NBTI and oxide degradation sensors with digital outputs. Due to their small size and simple frequency outputs they are amenable for use in standard cell-based designs. The sensors enable high-volume data collection and the monitoring of chip reliability throughout system lifetime. The degradation data supplied by these sensors also aids in understanding and modeling the complex degradation mechanisms. For example, results from a test chip in 130-nm CMOS provide new insight concerning the statistical nature of NBTI and gate oxide degradation, the impact of electric field on NBTI recovery, and the relationship between initial $V_{\mathrm{th}}$ and NBTI-induced $V_{\mathrm{th}}$ shift.

The oxide degradation sensor $(150\ \mu\mathrm{m}^2)$ monitors the change in gate leakage with degradation and is the first of its kind to be proposed. The NBTI sensor $(308\ \mu\mathrm{m}^2)$ is based on a subthreshold ring oscillator concept and is 110 smaller than previous work. We propose a simple calibration method to process the sensor output data. We observed a maximum error of 2.2% for the NBTI sensor under process/voltage/temperature variations, yielding $\Delta V_{\mathrm{th}}$ measurements with $3\sigma$ accuracy of 1.23 mV from 40 °C–110 °C.

## REFERENCES

[1] E. Karl, D. Blaauw, D. Sylvester, and T. Mudge, "Reliability modeling and management in dynamic microprocessor-based systems," in *Proc. ACM/IEEE Design Autom. Conf.*, 2006, pp. 1057–194.
[2] S. Y. Borkar, "Platform 2015: Intel processor and platform evolution for the next decade," Tech. Rep., Intel White Paper, 2005.
[3] J. Keane, T.-H. Kim, and C. H. Kim, "An on-chip NBTI sensor for measuring PMOS threshold voltage degradation," in *Proc. ACM/IEEE Int. Symp. Low Power Electron. Design*, 2007, pp. 189–194.
[4] T.-H. Kim, R. Persaud, and C. H. Kim, "Silicon odometer: An on-chip reliability monitor for measuring frequency degradation of digital circuits," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 874–880, Apr. 2008.
[5] E. Karl, D. Sylvester, and D. Blaauw, "Analysis of system-level reliability factors and implications on real-time monitoring methods for oxide breakdown device failures," in *Proc. IEEE Int. Symp. Quality Electron. Design*, 2008, pp. 391–395.
[6] E. Karl, P. Singh, D. Blaauw, and D. Sylvester, "Compact in situ sensors for monitoring NBTI and oxide degradation," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2008, pp. 410–623.
[7] D. K. Schroder, "Negative bias temperature instability: What do we understand?," *Microelectron. Reliab.*, vol. 47, no. 6, pp. 841–852, Jun. 2007.
[8] Huard, C. Parthasarathy, N. Rallet, C. Guerin, M. Mammase, D. Barge, and C. Ouvrard, "New characterization and modeling approach for NBTI degradation from transistor to product level," in *Proc. IEEE Int. Electron Devices Meet. (IEDM)*, 2007, pp. 797–800.
[9] B. Paul, K. Kang, H. Kufluoglu, M. A. Alam, and K. Roy, "Impact of NBTI on the temporal performance degradation of digital circuits," *IEEE Electron Device Lett.*, pp. 560–562, Aug. 2005.

[10] W. Wang, S. Yang, S. Bhardwaj, R. Vattikonda, S. Vrudhula, F. Liu, and Y. Cao, "The impact of NBTI on the performance of combinational and sequential circuits," in *Proc. ACM/IEEE Design Autom. Conf.*, 2007, pp. 364–369.
[11] S. Kumar, K. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," in *Proc. IEEE Int. Symp. Quality Electron. Design*, 2006, pp. 210–218.
[12] R. Vattikonda, W. Wang, and Y. Cao, "Modeling and minimization of PMOS NBTI effect for robust nanometer design," in *Proc. ACM/IEEE Design Autom. Conf.*, 2006, pp. 1047–1052.
[13] G. Chen, K. Y. Chuah, M. F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin, and D. L. Kwong, "Dynamic NBTI of PMOS transistors and its impact on device lifetime," *IEEE Electron Device Lett.*, pp. 734–736, Dec. 2002.
[14] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability," in *Proc. IEEE Int. Electron Devices Meet.*, 2003, pp. 341–344.
[15] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modeling," *Microelectron. Reliab.*, vol. 46, no. 6, pp. 1–23, 2006.
[16] M. Denais, C. Parthasarathy, G. Ribes, Y. Rey-Tauriac, N. Revil, A. Bravaix, V. Huard, and F. Perrier, "On-the-fly characterization of NBTI in ultra-thin gate oxide PMOSFET's," in *Proc. IEEE Int. Electron Devices Meet.*, 2004, pp. 109–112.
[17] S. Aota, S. Fujii, Z. W. Jin, Y. Ito, K. Utsumi, E. Morifuji, S. Yamada, F. Matsuoka, and T. Noguchi, "A new method for precise evaluation of dynamic recovery of negative bias temperature instability," in *Proc. IEEE Int. Conf. Microelectron. Test Structures*, 2005, pp. 197–199.
[18] C. Shen, C. E. Li, M.-F. Foo, T. Yang, D. M. Huang, A. Yap, G. S. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast Vth transient in NBTI of pMOSFETs with sion dielectrics," in *Proc. IEEE Int. Electron Devices Meet.*, 2006, pp. 1–4.
[19] R. Ferńandez, B. Kaczer, A. Nackaerts, S. Demuynck, R. Rodriguez, M. Nafria, and G. Groeseneken, "AC NBTI studied in the 1 Hz –2 GHz range on dedicated on-chip circuits," in *Proc. IEEE Int. Electron Devices Meet.*, 2006, pp. 337–340.
[20] J. Keane, D. Persaud, and C. H. Kim, "An all-in-one silicon Odometer for separately monitoring HCI, BTI, and TDDB," in *Proc. IEEE Symp. VLSI Circuits*, 2009, pp. 108–109.
[21] M. B. Ketchen, M. Bhushan, and R. Bolam, "Ring oscillator based test structure for NBTI analysis," in *Proc. IEEE Int. Conf. Microelectron. Test Structures*, 2007, pp. 42–47.
[22] R. H. Fowler and L. Nordheim, "Electron emission in intense electric-fields," *Proc. Royal Soc. London*, vol. 11, pp. 173–181, 1928.
[23] M. Lenzlinger and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown $SiO_2$," *J. Appl. Phys.*, vol. 40, pp. 278–283, 1969.
[24] I. C. Chen, S. E. Holland, and C. Hu, "A quantitative physical model for time-dependent breakdown in $SiO_2$," in *Proc. Int. Reliab. Phys. Symp.*, 1985, p. 24.
[25] Y. Uraoka, N. Tsutsu, Y. Nakata, and S. Akiyama, "Evaluation technology of VLSI reliability using hot carrier luminescence," *IEEE Trans. Semicond. Manuf.*, vol. 4, no. 3, pp. 183–192, 1991.
[26] S. R. Nariani and C. T. Gabriel, "A simple wafer-level measurement for predicting oxide reliability," *IEEE Electron Device Lett.*, vol. 16, pp. 242–244, 1995.
[27] M. Wang and K. Fu, "Statistical method of monitoring gate oxide layer yield," U.S. Patent 6 289 291, 2001.
[28] M. Acar, A.J. Annema, and B. Nauta, "Digital detection of oxide breakdown and life-time extension in submicron CMOS technology," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2008, pp. 530–633.
[29] 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.
[30] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature instability," *Appl. Phys. Lett.*, vol. 86, no. 14, pp. 1–3, 2005, Art. No. 142103.

**Prashant Singh** received the B.S.E. degree from the Indian Institute of Technology, Kanpur, India, in 2006, and the M.S.E. and Ph.D. degrees from University of Michigan, Ann Arbor, in 2011, all in electrical engineering.

He is currently a Design Engineer with Nvidia, Santa Clara, CA. His research interests include circuit reliability, low power design, power integrity.

**Eric Karl** (S'03–M'08) received the B.S.E., M.S.E., and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, in 2002, 2004, and 2008, respectively.

In 2001, he was an intern with SPARC Microprocessor Development Team, Sun Microsystems. In 2004, he was a graduate intern with the IBM T. J. Watson Research Center and in 2005, he was a graduate intern with the Circuits Research Laboratory, Intel Corporation. After finishing the Ph.D. in March 2008, he joined Logic Technology Development, Intel Corporation, Hillsboro, OR, as a Senior Design Engineer. Since then, he has been working on low-power, high-performance SRAM cache design and technology development for CPU and SoC applications. He is the author or coauthor of 13 technical papers.

Dr. Karl participated in reviewing papers for conferences and journals including the International Symposium on Computer Architecture, Solid State Electronics, ACM/IEEE Design Automation Conference, and IEEE TRANSACTIONS ON ELECTRON DEVICES.

**David Blaauw** received the B.S. degree in physics and computer science from Duke University, Durham, NC, in 1986, and the Ph.D. degree in computer science from the University of Illinois, Urbana, in 1991.

Until August 2001, he worked for Motorola, Inc., Austin, TX, where he was the manager of the High Performance Design Technology Group. Since August 2001, he has been on the faculty at the University of Michigan where he is a Professor. He has published over 350 papers and hold 40 patents. His work has focussed on VLSI design with particular emphasis on ultra low power and high performance design.

Prof. Blaauw was the Technical Program Chair and General Chair for the International Symposium on Low Power Electronic and Design. He was also the Technical Program Co-Chair of the ACM/IEEE Design Automation Conference and a member of the ISSCC Technical Program Committee.

**Dennis Sylvester** (S'95–M'00–SM'04–F'11) received the Ph.D. degree in electrical engineering from the University of California, Berkeley, where his dissertation was recognized with the David J. Sakrison Memorial Prize as the most outstanding research in the UC-Berkeley EECS Department.

He is a Professor with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, and Director of the Michigan Integrated Circuits Laboratory (MICL), a group of ten faculty and 60+ graduate students. He previously held research staff positions with the Advanced Technology Group of Synopsys, Mountain View, CA, Hewlett-Packard Laboratories, Palo Alto, CA, and a visiting professorship with the Department of Electrical and Computer Engineering, National University of Singapore. He has published over 300 articles along with one book and several book chapters. His research interests include the design of millimeter-scale computing systems and energy efficient near-threshold computing for a range of applications. He holds 7 U.S. patents. He also serves as a consultant and technical advisory board member for electronic design automation and semiconductor firms in these areas. He co-founded Ambiq Micro, a fabless semiconductor company developing ultra-low power mixed-signal solutions for compact wireless devices.

Dr. Sylvester was a recipient of an NSF CAREER Award, the Beatrice Winner Award at ISSCC, an IBM Faculty Award, an SRC Inventor Recognition Award, and eight Best Paper Awards and Nominations. He was the recipient of the ACM SIGDA Outstanding New Faculty Award and the University of Michigan Henry Russel Award for distinguished scholarship. He has served on the technical program committee of major design automation and circuit design conferences, the executive committee of the ACM/IEEE Design Automation Conference, and the steering committee of the ACM/IEEE International Symposium on Physical Design. He is currently an Associate Editor for the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS and previously served as an Associate Editor for the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS. He is a member of ACM and Eta Kappa Nu.