An Array-Based Odometer System for Statistically Significant Circuit Aging Characterization

John Keane, *Member, IEEE*, Wei Zhang, *Student Member, IEEE*, and Chris H. Kim, *Senior Member, IEEE*

***Abstract*—Variations in the number and characteristics of charges or traps contributing to transistor degradation lead to a distribution of device "ages" at any given time. This issue is well understood in the study of time dependent dielectric breakdown, but is just beginning to be thoroughly addressed under bias temperature instability (BTI) and hot carrier injection (HCI) stress. In this paper, we present a measurement system that facilitates efficient statistical aging measurements involving the latter two mechanisms in an array of ring oscillators. Microsecond measurements for minimal BTI recovery, as well as frequency shift measurement resolution ranging down to the error floor of 0.07% are achieved with three beat frequency detection systems working in tandem. Measurement results from a 65 nm test chip show that fresh frequency and the stress-induced shift are uncorrelated, both the mean and standard deviation of that shift increase with stress, and the standard deviation/mean ratio decreases with stress time.**

***Index Terms*—Aging, bias temperature instability, circuit reliability, digital measurements, hot carriers.**

I. Introduction

TRANSISTOR aging is the product of a finite number of trapped charges or broken bonds in aggressively scaled modern devices, so it is no longer sufficient to rely on a small set of stress tests to predict the behavior of up to billions of devices over the lifetime of a circuit. Statistical fluctuations in the number and spatial distribution of defects contributing to transistor degradation lead to a range of effective device "ages" at any given time [1]–[13]. This issue is well understood in the study of time dependent dielectric breakdown (TDDB), but has yet to be fully addressed under bias temperature instability (BTI) and hot carrier injection (HCI) stress. The circuit conditions leading to those degradation processes are illustrated in Fig. 1.

NBTI (Negative Bias Temperature Instability) in PMOS is characterized by a positive shift in the absolute value of the V_{th} (threshold voltage), which occurs when a device is biased in strong inversion with a small, or no lateral electric field (i.e., $V_{DS} \approx 0$ V). The V_{th} shift is generally attributed to hole trapping in the dielectric bulk, and/or to the breaking of Si-H bonds at the gate dielectric interface by holes in the inversion layer, which generates positively charged interface traps. When a stressed device is turned off, it immediately enters the "recovery" phase, where trapped holes are released, and/or the freed hydrogen species diffuse back towards the substrate/dielectric interface to anneal the broken Si-H bonds, thereby reducing the absolute value of the V_{th}. Positive bias temperature instability (PBTI) in NMOS transistors was not critical in silicon dioxide dielectrics (such as those used in the current test circuit implementation), but is now contributing to the aging of high-k gate stacks [2].

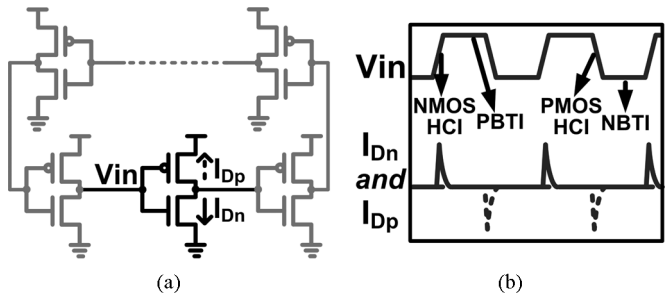


Fig. 1. (a) A standard inverter-based ROSC. (b) Associated voltage and current waveforms demonstrating the combination of BTI ($V_{DS} = 0$ V) and HCI ($V_{DS} \neq 0$ V) stress that devices undergo when the ROSC is oscillating.

HCI has become less prominent with the reduction of operating voltages, but remains a serious concern due to the large local electric fields in scaled devices. Hot carriers (i.e., those with high kinetic energy) accelerated toward the drain by a lateral electric field across the channel lead to secondary carriers generated through impact ionization. Either the primary or secondary carriers can gain enough energy to be injected into the gate stack. This creates traps at the silicon substrate/gate dielectric interface, as well as dielectric bulk traps, and hence degrades device characteristics such as the V_{th}.

The number of defects contributing to aging is reduced in smaller transistors, making the relative impact of the creation or destruction of each of them more significant. Much like variations caused by random dopant fluctuations, aging-induced variation scales inversely with gate area [1], [2]. In addition to variations in the number of defects, the size of the step they each cause in the measured parameter of interest (e.g., V_{th}), is also randomly distributed. That step size was reported to be dependent on each charge's position relative to the others [3]. Recent work has shown that a single trapping/detrapping event in a pre-existing oxide trap can lead to a V_{th} shift of more than 30 mV in scaled PMOS devices with high-k gate dielectrics, and these step heights are exponentially distributed [4]. Several authors

Manuscript received March 14, 2011; revised May 26, 2011; accepted June 06, 2011. Date of publication July 22, 2011; date of current version September 30, 2011. This paper was approved by Associate Editor Stefan Rusu. This work was supported in part by the SRC under award 2008-HJ-1805, and in part by Broadcom and the IBM PhD Fellowship program.

J. Keane was with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA, and is now with Intel Corporation, Hillsboro, OR 97124 USA (e-mail: john.keane@intel.com).

W. Zhang and C. H. Kin are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

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/JSSC.2011.2160813

have stated that the widely distributed time constants of the defects involved in BTI stress contribute to the variation [4]–[7].

Most of the research in this area has focused on device-level measurements, and efforts to model the impact of the spread in aging characteristics on analog applications or SRAM stability. Little has been done to investigate the effects of varied aging characteristics in digital logic. In this paper we present a measurement system that facilitates statistical aging measurements involving HCI and BTI in ring oscillators (ROSCs), although more emphasis will be placed on the latter. The distribution of frequencies is monitored by three Silicon Odometer beat frequency detection systems working in parallel [15]–[17]. This system avoids errors caused by neglecting unwanted BTI recovery with measurement times of down to 1 μs. Frequency shift measurement resolution ranging down to the error floor of 0.07% is achieved along with those quick measurements, which is typically not possible in standard test setups. Hofmann *et al.* recently described a product-level aging monitoring system where measurement times of 80 μs were achieved with 0.2% resolution [18]. However, no details were given about how frequencies were recorded in this single-ended system. We previously compared standard on-chip frequency measurement systems, showing that the two main benefits of the Odometer are its differential nature, and the fact that its resolution is set by the difference between two frequencies rather than the ROSC under test frequency itself [16].

II. Prior Work in Statistical Aging Characterization

It has long been understood that TDDB is governed in part by a statistical component related to the density of defects required to form a conducting path through the gate dielectric [19]. Until recently BTI and HCI have not been found to display significant spreads in their degradation characteristics. However, this has quickly changed as device widths and lengths scale down to the dimensions that gate thicknesses achieved long ago. Rauch demonstrated that identical PMOS devices stressed under the same conditions will experience different amounts of degradation, and that this variance increases with stress time [8]. He also observed that the mismatch shifts in paired PMOS were not correlated to the initial mismatch values.

La Rosa *et al.* investigated the impact of variations in NBTI on SRAM cells [9]. Rauch then summarized both his mismatch work and this SRAM study in a 2007 paper before moving on to define the shape of the aging-induced threshold shift distributions [3]. He stated that NBTI degradation is the net result of two Poisson processes: the creation and destruction of charges. The difference between two Poissons is called a Skellam distribution, and measured results were shown to closely fit this model. Both the mean and variance of the threshold shift ($\mu(\Delta V_{th})$ and $\sigma^2(\Delta V_{th})$, respectively) were found to increase with stress time, but $\sigma^2(\Delta V_{th})/\mu(\Delta V_{th})$ decreases.

Huard *et al.* also investigated the impact of NBTI variation on SRAM arrays [10]. The authors found that $\sigma^2(\Delta V_{th})$ is linearly proportional to $\mu(\Delta V_{th})$, and that data from large sample sets fit Skellam distributions. Huard and Rauch claimed that the non-normality of the ΔV_{th} distribution is smoothed out to a degree in large SRAM arrays when it is convolved with the fresh Gaussian V_{th} distribution. Fischer *et al.* built a dense test structure in 65 nm technology to mimic the stress that PMOS devices undergo in a standard 6T SRAM array [11]. They found that while V_{th} and I_D distributions are Gaussian before and after stress, the ΔV_{th} is not Gaussian. However, the authors observed that the Skellam underestimated the high tail of their ΔV_{th} distribution, and speculated that the discrepancy with Huard's work may be due to different measurement techniques. Kaczer *et al.* presented an analytical model for the NBTI ΔV_{th} distribution based on the finding that the charging and discharging of individual oxide defects during stress and recovery result in exponentially distributed V_{th} shifts (linking it to RTN), and the often used argument that the number of charged defects per device is Poisson-distributed [4].

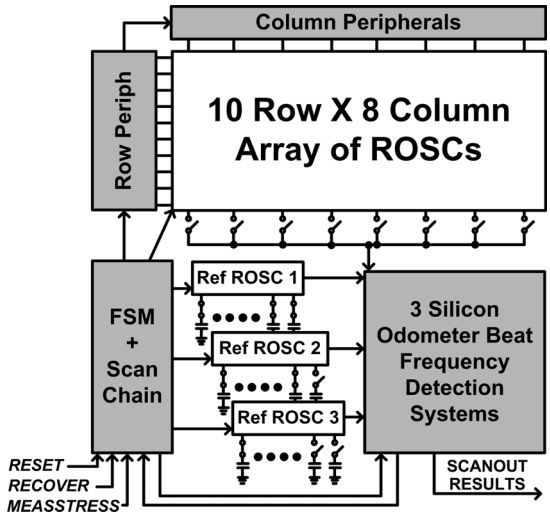


Fig. 2. Top level system diagram. The reference ROSCs each have 15 trimming capacitors controlled by the scan chain. We were limited to 80 ROSC cells in the array due to silicon area constraints, but this design could be expanded to include many more cells.

Kang *et al.* presented a statistical NBTI model based on the Reaction Diffusion (RD) framework [12]. The authors claim that $\sigma(V_{th})$ increases with stress time, having a power law exponent that is half of that followed by the mean. Pae *et al.* presented BTI variation measurements from 90 nm, 65 nm, and high-k, metal gate 45 nm technologies [2]. Their data indicates that the initial threshold voltage spread is not correlated to the BTI-induced variability, and $\sigma(V_{th})$ increases with stress time following power law behavior. Finally, Wang *et al.* developed a statistical methodology to predict circuit performance under the influence of process variations and NBTI [20]. The authors state that the standard deviation in logic speed decreases with stress time, following a power law exponent of 1/6 regardless of the amount or type of variations (local or global). This is contrary to other studies of variation in NBTI summarized in this section, as well as our own measurement results, which will be presented in Section IV-C.

III. Statistical Odometer System Design

Our system consists of a 10×8 array of cells containing ROSCs to be stressed, a finite state machine (FSM), a scan chain, and three Silicon Odometers with their reference ROSCs

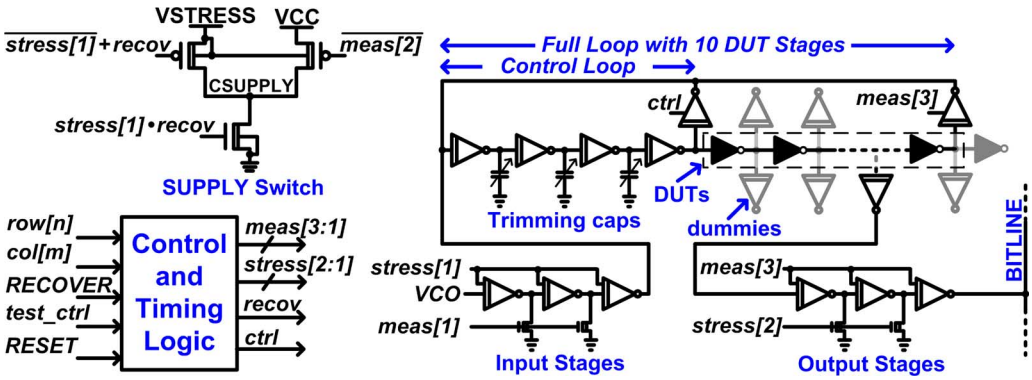


Fig. 3. ROSC cell design. The thin oxide logic stages under test are colored black, and all other transistors are thick oxide I/O devices (indicated by double lines).

(Fig. 2). We were limited to 80 ROSC cells due to the available silicon area, but this design could be expanded. During tests the whole array of ROSCs, or any one rectangular group of them, are stressed in parallel and selected one-by-one for measurements. Alternatively, all of the ROSCs under test can be put into a recovery state (i.e., 0 V supply), along with any cells that are not selected. During stress, the ROSC loops are opened so that their frequencies can be controlled by an on-chip voltage controlled oscillator (VCO). When each oscillator is selected for a measurement with the *MEASSTRESS* signal from the controlling software, its supply is set to the standard digital level of 1.2 V, the loop is closed, and its frequency shift is measured by the three Odometer systems. Reference ROSCs are put into a 0 V no-stress state in between measurements, and their supply is set at 1.2 V during those brief events so they should not age appreciably. Even if short periods under the nominal supply are sufficient to cause any aging, the effect on our results is cancelled out by the differential measurement setup.

A. Ring Oscillator Cell Design

Each ROSC cell contains its own supply switch that sets the local virtual supply (CSUPPLY) at the stress level (VSTRESS), 1.2 V (VCC), or 0 V as appropriate (Fig. 3). The ROSC cells also include selection logic to switch a cell into measurement mode if its $row[n]$ and $col[m]$ signals are both high, by closing the loop and then connecting one tapped output node to the bitline after the virtual supply has had time to settle to VCC. This timing is indicated in Fig. 3 by the order of the *meas* ("measure") and *stress* signals. The former go high during measurement periods, and the latter are driven high during stress or recovery. First, *meas[1]*/*stress[1]* turn off the input stages and the VSTRESS switch. Next *meas[2]* turns the VCC switch on, and finally *meas[3]*/*stress[2]* close the loop and connect it to the bitline.

When a cell is sent back to stress, or recovery if the *RECOVER* signal is high, this ordering is roughly reversed. The ROSC is first disconnected from the bitline to prevent any unwanted stress in other parts of the array. At nearly the same time, the ROSC is opened, the VCC switch is turned off, and the input path from the VCO is turned on. Finally, the VSTRESS or GND supply switch is closed to start stress or recovery, respectively. Several tri-state inverters and pull-down transistors are placed between the VCO input and the ROSC, as well as the ROSC and the bitline, in order to prevent any coupling when those connection paths are shut off. The *RESET* signal asynchronously sends the whole array into recovery mode.

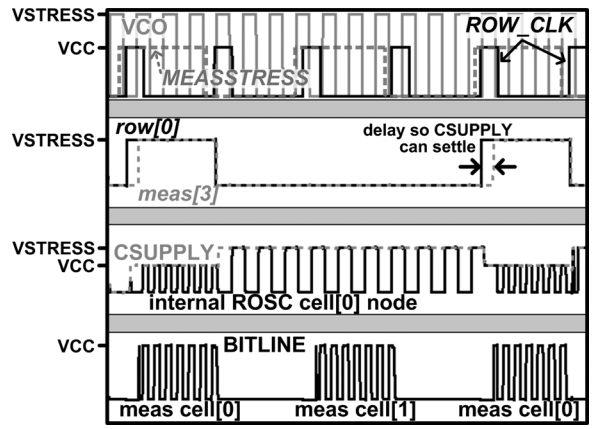


Fig. 4. Waveforms illustrating the basic operation of a ROSC cell. In this simulation, only two cells are included in order to demonstrate the functionality as cell[0] goes into and out of stress periods.

Basic cell operation is illustrated in Fig. 4. Only two cells are included in this simulation to demonstrate the functionality of one of them as it enters and leaves stress mode. The external *MEASSTRESS* signal controls the internal *ROW_CLK* signal, which starts and then stops each measurement with two consecutive pulses, as will be explained in Section III-C. The delay between CSUPPLY dropping to VCC and *meas[3]* closing the loop ($\sim$ 7 ns under nominal conditions) prevents unstable oscillations. Ten inverters in each ROSC are constructed with 1.2 V thin oxide logic devices under test (DUTs). These stages will age during stress experiments while the rest of the gates, composed of 2.5 V thick oxide I/O transistors, do not exhibit any noticeable degradation. However, the thick oxide control stages contribute to the full loop delay, which must be accounted for when calculating the stressed stages' delay shift due to aging.

Therefore, in addition to the full loop that is selected with the *meas[3]* signal, we added a replica control path selected with *ctrl* which is asserted at the appropriate time if the scan bit *test_ctrl* is high. The delay of the replica path is roughly equivalent to that of the control logic in the full loop. So by first measuring the control loop frequency, and then that of the full loop during automated circuit calibration, we can calculate

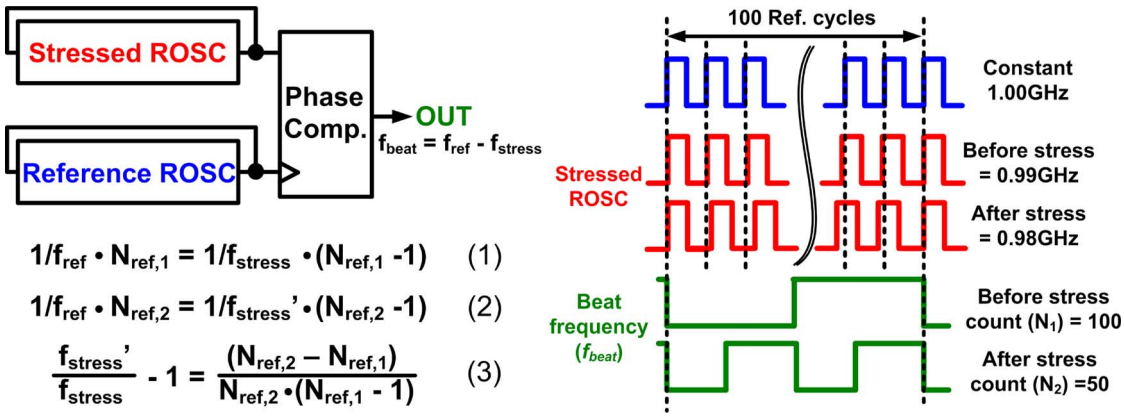


Fig. 5. Concept behind the beat frequency detection system. (1) is valid before stress, and the $(N_{ref,1} - 1)$ term arises from the fact that the slower stressed ROSC will take one less period to cycle back to the same point in the reference ROSC period while both are oscillating. After stress, only the count and f_{stress} change (2). We can calculate the stressed ROSC frequency shift with these two equations as shown in (3).

the percentage of the fresh full loop delay accounted for by the DUTs (i.e., $1 - f_{full}/f_{ctrl}$). Extracted simulations showed the real delay of the DUT stages to be only 0.55% longer than the calculated value. During experiments the total frequency shift of each stressed full loop is divided by the percentage of the fresh delay taken by the DUTs, in order to calculate the degradation in those thin oxide stages. All DUT stages have identical loads and layouts due to the use of dummy cells.

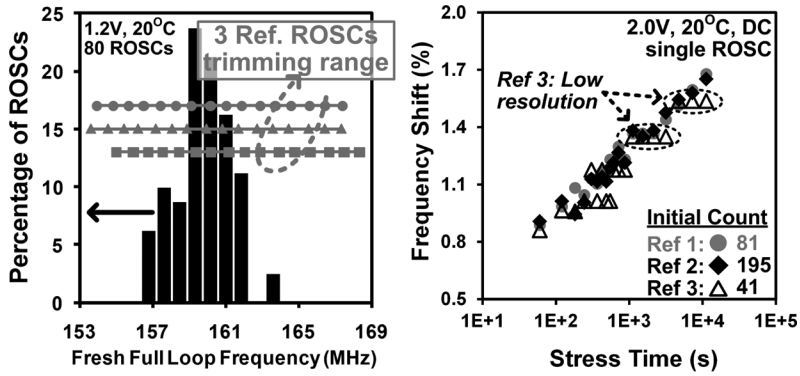
B. Multiple Silicon Odometer Beat Frequency Detection Setup

The Silicon Odometer measures frequency changes in the stressed ROSCs with the concept illustrated in Fig. 5. During the short measurement periods, a phase comparator uses a fresh reference ROSC to sample the output of an identical stressed ROSC. The output signal of this phase comparator exhibits the beat frequency: $f_{beat} = f_{ref} - f_{stress}$. A counter is used to measure the beat frequency by counting the number of reference ROSC periods during one period of the phase comparator output signal. This count is recorded after each stress period to calculate the shift down in f_{stress}. The Odometer measures frequency changes with resolution and measurement times theoretically ranging down to $< 0.01\%$ and $< 1\ \mu$s respectively [15], [16].

This differential system provides high-resolution frequency shift measurements when the frequencies of the ROSC under test and reference are close. That is because the beat frequency of two input signals that are very close in frequency is low as shown in the f_{beat} equation above, meaning we will count more periods of the reference ROSC during one beat period, and a small change in f_{stress} then causes a larger change in the count. This leads to longer stress interruptions because more periods of the constant f_{ref} equates to more measurement time, but there is still a clear resolution versus measurement time benefit compared with simple ROSC period counters (see [16]). A small initial difference between f_{ref} and f_{stress} is ensured with trimming capacitors, and in past Odometer test circuits each capacitor on both oscillators has been individually controlled with scan chain bits. However, in the present circuit where many ROSCs are stressed in parallel and selected one-by-one for measurements, controlling the capacitors in each stressed oscillator would be time and area consuming. Therefore, we instead hardwired nine of fifteen capacitors "on" in each of those ROSCs, while individually controlling all fifteen in the three references.

The Odometers associated with those references all record counts corresponding to the beat frequency for each ROSC measurement. During post-processing, the highest-resolution degradation characteristic is selected from that set of three for each oscillator that was stressed. Fig. 6(a) presents an example distribution of 80 fresh full loop frequencies, along with the range covered by the three reference ROSCs under all trimming conditions. Turning on each trimming capacitor slowed a reference ROSC by roughly 900 kHz, or 0.57% of the mean fresh full loop frequency under nominal operating conditions. During calibration, the reference ROSCs are trimmed to positions within the fresh array distribution such that we maximize the resolution of the group of degradation characteristics gathered from each full stress experiment.

Fig. 6(b) shows a group of three degradation characteristics gathered by the references from one ROSC under test. Starting measurements with the reference and test ROSC frequencies close together, but the latter slightly slower, leads to a high resolution measurement with a monotonic decrease in the output counts [16]. The odometer output count is equal to the number of reference ROSC periods (N_{ref}) counted throughout one period of the beat frequency, during which time one less cycle is observed in the stressed ROSC ($N_{stressed} = N_{ref} - 1$). Therefore, according to (1) from Fig. 5 with $N_{ref} = 41$, reference ROSC 3 started out 2.44% faster than the ROSC under test in the current example (i.e., $1 - N_{stressed}/N_{ref} = 1 - 40/41$). The resulting low resolution is apparent from the highly quantized outputs of the corresponding Odometer. However, reference 2 was initially only 0.513% faster than the ROSC under test, so we can select this high resolution result for our analysis instead. Finally, note that we can also use measurement results from a reference ROSC that starts out *slower* than the ROSC under test, so the output count will *increase* with stress [16]. Therefore, we can more easily cover the distribution of fresh frequencies in our array with only three appropriately trimmed references.



(a) (b)

Fig. 6. (a) Example of fresh full loop frequency distribution for 80 cell array, with the corresponding reference ROSC trimming range. (b) Measured results from all three odometers for one ROSC under test. The ROSC under test started out slower than all references in this case.

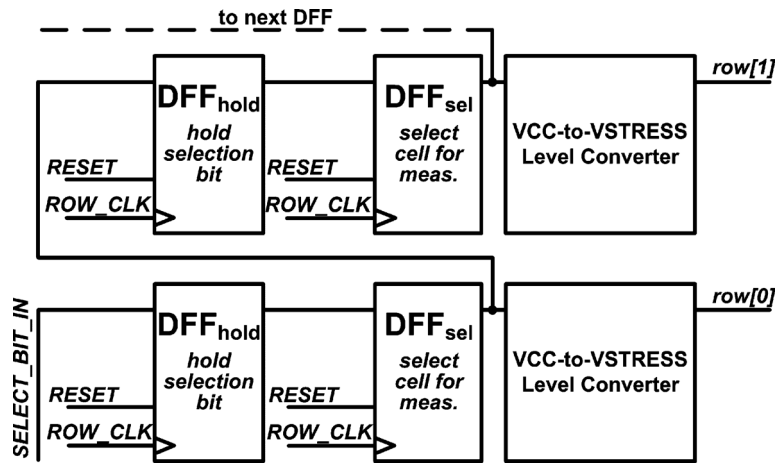


Fig. 7. Row selection logic ("Row Periph" from Fig. 2). Two flops are included for each row so that as soon as a measurement is finished on $row[n]$, the selection bit can move to $row[n+1]$ without starting a measurement there until the system is ready.

C. Test Interface and Procedure

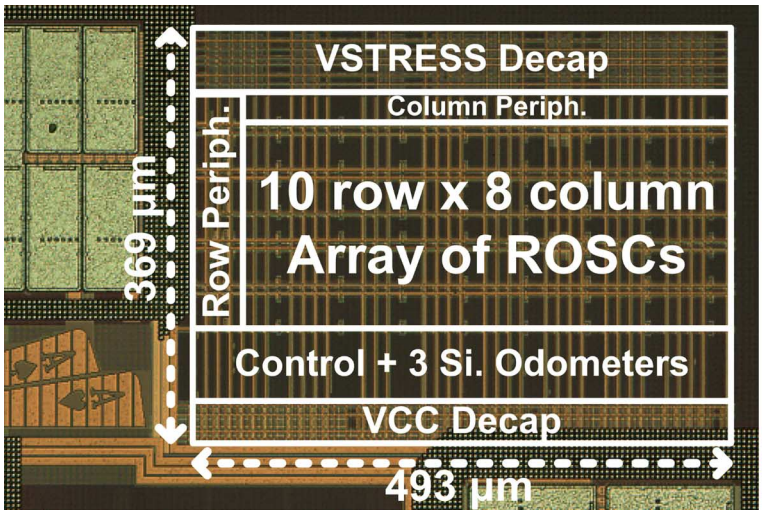
Calibration and measurements are automated through a simple digital interface. During calibration we record the fresh control and full loop frequencies from each ROSC in the array by reading those values with an oscilloscope. The error in this step is minimized by averaging thirty results for each loop during the automated measurement routine. Next, we sweep through the trimming range in the three reference ROSCs, again averaging the measured frequency results from thirty samples. After that point, the optimal trimming configurations are selected in order to cover the distribution of frequencies of the ROSCs to be tested.

A *RESET* signal is asserted before stress conditions are set which prevents any cells from being selected, and puts them all into recovery mode. During experiments, ROSC cells are cycled through without the need to send or decode cell addresses, in order to simplify the logic and attain faster measurement times. The first cell is selected with an initialization sequence, and *MEASSTRESS* is asserted each time that the controlling software is ready for a new measurement. The row selection signal is incremented with each measurement, and the column selection shifts after all of the cells in a column have been selected. Any cells not selected for stress are kept in a 0 V no-stress state by asserting the *RECOVER* signal appropriately.

The logic used to store the row selection signal had to minimize the time when a ROSC is taken out of stress in order to prevent unwanted BTI recovery. Although this Odometer provides measurement times of down to 1 μs, we still have to account for the time it takes to scan out the results. Therefore, two DFFs were used for each row, as shown in Fig. 7. *SELECT_BIT_IN* is clocked into the first $\mathrm{DFF_{hold}}$ with *ROW_CLK* during initialization. The next *ROW_CLK* pulse starts a measurement on *row[0]* by moving the select bit to the $\mathrm{DFF_{sel}}$ on the first rising edge of *MEASSTRESS*. That selection bit is then sent to the $\mathrm{DFF_{hold}}$ in *row[1]* by another pulse of *ROW_CLK* on the falling edge of *MEASSTRESS*, and is held there while the results are scanned out. The next assertion of *MEASSTRESS* from the controlling software starts a measurement on *row[1]*, and this process is repeated through the rows and columns as necessary.

IV. Statistical Odometer Test Chip Measurements

A 369×493 $\mu\mathrm{m}^2$ test circuit was implemented in 65 nm bulk CMOS. Fig. 8 presents a die photo and the test chip characteristics. Measurements were automated with LabVIEW™ software through a National Instruments data acquisition board. The



Process	65nm LP CMOS, 7M
Logic / I/O supplies	1.2V / 2.5V
Active Area	~257x475µm^2
Total Area	369x493µm^2
Odom. Δf Error Floor	0.07%
Measure Interrupt	≥1µs
DUT dimensions	P: 300/60nm N: 200/60nm
w/in die 10 stage DUT freq. σ/µ	1.32% - 1.78%

Fig. 8. Die photo and summary of the test chip characteristics.

DUTs were 200 nm/60 nm NMOS and 300 nm/60 nm PMOS transistors in the inverter stages of the stressed ROSCs. While much of the data from previous work outlined in Section II was gathered from single device measurements, our system stresses half of the NMOS and half of the PMOS in each ROSC under test at any given time, which represents real stress conditions in many digital circuits.

PBTI in the NMOS is not significant when high-k dielectrics are not used, so this means that only half of the PMOS devices and none of the NMOS undergo substantial BTI aging in between transitions in this system. However, when the ROSC begins oscillating in closed-loop mode during measurements, both the degraded and the relatively unstressed paths are exercised. Under AC stress all PMOS are stressed for half of the time between transitions, and HCI degradation occurs in both types of transistors during the switching events. In either case since the drive current, and therefore the frequency, is now linearly proportional to V_{th}, we can compare our results to the trends found in the studies of V_{th} degradation [18], [21].

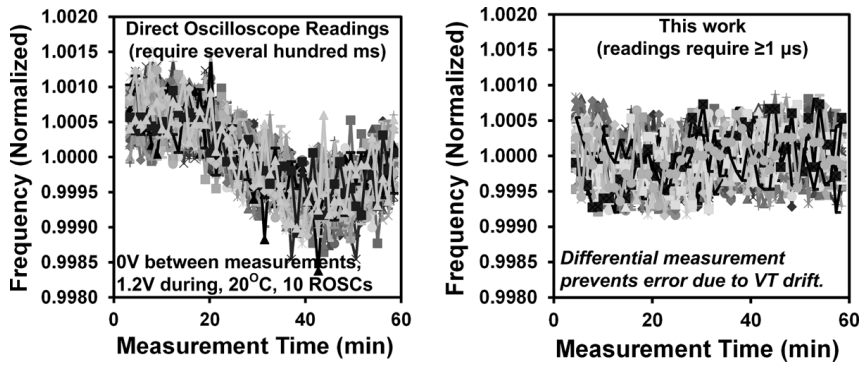
Measurement times were set to 2.5 μs unless noted otherwise, which allowed most beat frequency counts to complete correctly. As discussed in a previous publication, the initial two counts reported by the Odometer are unpredictable because they begin at random mid-points in the beat frequency period [16]. Therefore, we record the last three count results that are latched before the measurement period is stopped, and use the final count which can be verified to be correct during calibration. If the time required to latch these first three counts is greater than the chosen measurement time due to a long beat frequency period (i.e., high Odometer output count), or long reference and stressed ROSC periods, then the desired results cannot be recorded. Shorter measurements were possible, but in that case the higher count results which did not have time to complete were discarded, and the next-highest resolution output was selected for post-processing. In addition to allowing more beat frequency counts to complete, we maintained 2.5 μs interrupts because measurement results showed that the difference between the frequency degradation measured with this value and 1 μs was negligible in our system (see Section IV-B).

A. Measurement Error Characterization

Fig. 9 illustrates frequency measurement results from 0 V, no-stress experiments, so ideally there should be no shift (i.e., the normalized frequency should remain at 1.0). The characteristics of forty ROSCs are displayed, and are representative of results seen from the entire arrays measured on multiple chips. Fig. 9(a) was directly recorded by a 100 MHz, 1.25 GS/s oscilloscope, after the frequencies were divided down by 1024 on-chip. We see a worst case error of 0.18%, and a drift in the measured values due to some slight change in operating conditions. Fig. 9(b) shows a smaller worst case error of 0.07% in the frequency calculated by the Silicon Odometer, along with the fact that this differential system eliminates the effects of variations common to the reference ROSC and that under test. Similar error floors were found for these systems during repeated tests, setting the lower bound on the range of frequency shifts that they can accurately measure. Finally, note that the automated oscilloscope readings required over 500 ms, while the Odometer measurements take ≥ 1 μs. This combination of high resolution and fast measurements is critical when measuring BTI stress, where we must avoid recovery when stress conditions are temporarily removed for readings.

B. Impact of Measurement Time on Bti Results

Fig. 10 presents ROSC frequency degradation results from DC stress conditions using a range of measurement times. Fig. 11 summarizes the findings from those measurements and illustrates the impact of interruption times on BTI measurements. Long stress interruptions, during which we record our measurements, take up a significant portion of the total experiment time at early measurement points. This means a large percentage of the time is spent in recovery state, which pulls down the early results and leads to a steeper degradation slope, as seen in Fig. 11(a). Several previous publications have clearly demonstrated this phenomenon [22]–[26]. Fig. 11(b) shows the power law exponents that were fit to the data from different measurement times. As times get into the millisecond range, the average exponent approaches 0.165, which has commonly



(a) (b)

Fig. 9. Error (i.e., deviation from 1.0) in (a) oscilloscope and (b) faster Odometer measurements from 40 ROSCs during no-stress experiments.

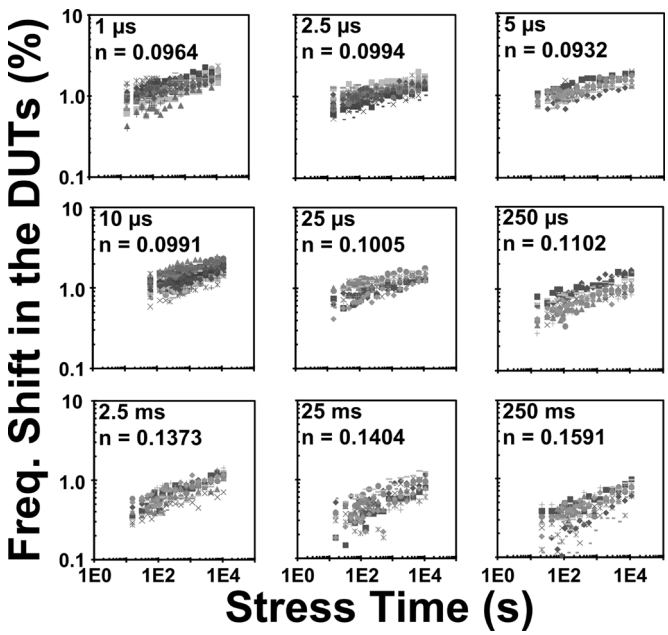
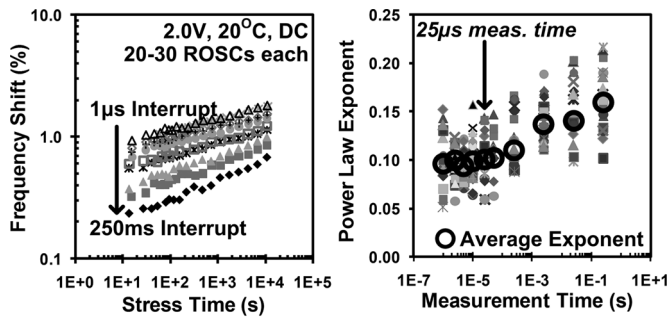


Fig. 10. Frequency degradation curves gathered at 2.0 V, 20° C, DC conditions with a range of stress interruption times for measurements. Each plot shows results from 20–30 ROSCs, and the increasing average power law exponent is specified.



(a) (b)

Fig. 11. (a) Steeper slopes for longer stress interruptions due to recovery. (b) Power law exponent versus measurement time. Measurement times (i.e., stress interruptions) of roughly 25 μs or less are sufficient to prevent measureable unwanted recovery in this system.

been cited as the measured and theoretically correct value. It fits the RD model theory if interface trap creation and passivation is assumed to be the sole cause of the BTI transient behavior, and H_2 is the hydrogen species diffusing through the dielectric [25]. However, our results, along with those from several other fast measurement techniques, show that this model and theory seem to be incomplete [7], [22], [25], [26].

The discrepancy with that traditional RD theory has sometimes been explained by claiming that NBTI appeared to be the result of two component processes: (1) a slower creation of interface traps (i.e., donor-like states resulting from broken Si-H bonds at the Si-SiO_2 interface) and subsequent diffusion of H_2 into the gate dielectric, along with (2) a faster hole trapping/detrapping process in the dielectric bulk [14], [22], [25], [28]–[31]. The former recovers on the scale of tens of microseconds or more, and the latter begins to recover within several microseconds or less. Therefore, slow measurement techniques requiring milliseconds or seconds miss much of the fast hole detrapping process, so the observed results are primarily due to the slower interface trap creation mechanism. This process has been claimed to have a higher power law exponent which lies in the 0.165 range, while fast hole trapping has a smaller exponent ($\leq$ 0.1). Recent work by Reisinger and others has called into question this distinction between "slow" and "fast" BTI components, and demonstrates a spectrum of trapping/detrapping times ranging down to the microsecond range or lower [27]. The dominant mechanism behind NBTI is still disputed, as other researchers state that it is primarily due to interface trap generation and annealing, so the RD model can describe at least this critical portion of the aging process with a high degree of accuracy [30], [31]. In any case, fast measurement techniques are required to directly observe the full range of NBTI degradation and recovery characteristics.

Our measurements show that stress interruptions of tens of microseconds or less are required to observe the average power law exponent of $\sim$ 0.1 under the listed stress conditions (Fig. 11). Ji *et al.* found that measurement times of $\leq$ 40 μs did not show any measureable recovery [24], although many other authors have claimed that this process can be recorded down to 1 μs or less [22], [25], [32]. The lack of measureable recovery in our setup below $\sim$ 10 μs could be due to secondary circuit effects and an insufficiently large sample set. However, the spread in the range of exponents observed at each measurement

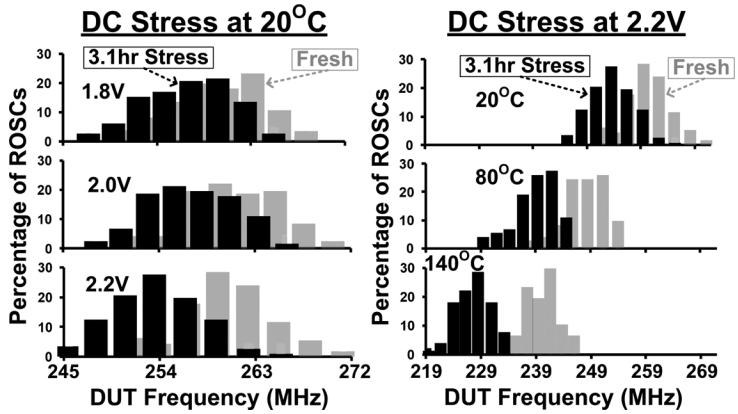
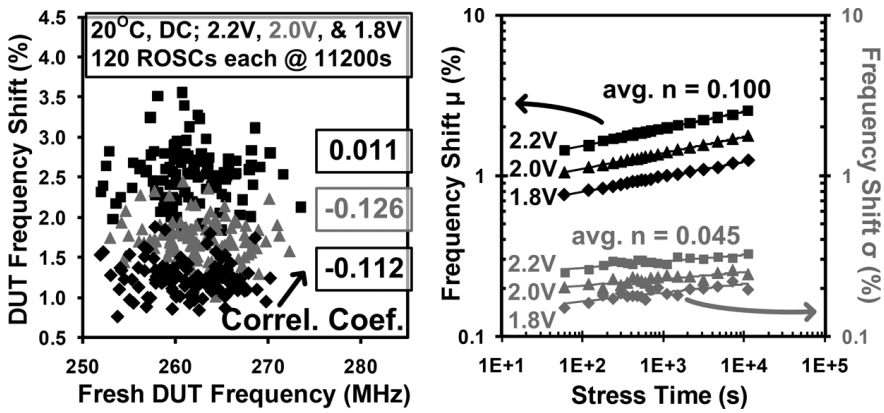


Fig. 12. Frequency distributions after 3.1 hour stress (black bars), along with the fresh distributions (gray bars). Each 20°C distribution was gathered from 120 ROSCs, while those at higher temperatures came from 80 due to a limited amount of dies.



(a) (b)

Fig. 13. (a) The fresh ROSC frequencies were uncorrelated with the stress-induced shifts. (b) The mean and standard deviation of the frequency shifts both increase with stress time.

time (roughly $+/-0.05$) illustrates the variation in the aging process, and the importance of characterizing a statistically significant sample set.

C. DC Stress Results

Histograms of fresh DUT frequencies are shown in Fig. 12 with the resulting distributions after 3.1 hours of DC stress (11,200 s). The distributions at 20°C were obtained from 120 ROSCs each, while the higher temperature measurements involved 80 ROSCs. Smaller numbers were used when necessary due to the limited number of dies we had to work with. In order to prevent any particular systematic inter-die process shifts from fully impacting one set of the results, each experiment that spanned more than 40 ROSCs came from multiple dies. For example, each of the 120 ROSC experiments came from three different dies. The primary degradation mechanism at work in these DC experiments was NBTI, and in Fig. 13(a) we see that there was no correlation between the fresh ROSC frequency and the stress-induced shift. This lines up with previous findings that the stress-induced V_{th} mismatch in PMOS pairs was uncorrelated to the initial mismatch [8], and that the initial spread in the V_{th} is not correlated to that caused by aging [2], [4], [11]. Fig. 13(b) shows the average (μ) frequency shifts and the standard deviation (σ) of the shifts versus stress time. The σ increases with stress [1], [2], [10], [12], roughly following a power law with an exponent (n) of just under 1/2 that of the μ shift. Therefore, the σ/μ ratio of the shift decreases with stress time [3].

The σ of the frequency did not show a clear trend with stress time. This value was poorly fitted by the power law (R-squared values of only $\sim 0.03 - 0.30$), with the exponent of this fit ranging from -0.002 to 0.028, meaning the σ value remained generally flat during stress. That behavior is expected because the spread in the fresh frequency is larger than that of the spread in the aging-induced shifts, and the increase in that latter value is modest during stress, as seen in Fig. 13(b). This trend stands in contrast to results from previous work that found a decrease in the σ of a path delay with stress time showing a power law exponent of 1/6 regardless of the type of variation [20]. It is possible that a systematic process shift could cause circuit speed variation to decrease with aging (e.g., a shorter gate length could lead to faster gates and worse HCI degradation), but in our measurements no location-dependent shifts were found in the aging data.

In Fig. 14, we see that the measured frequency shift distributions fit well to a lognormal distribution. Rauch found that the lognormal overestimated the high tail of his measured ΔV_{th} shifts, and that his data fit well to a Skellam with tails in between

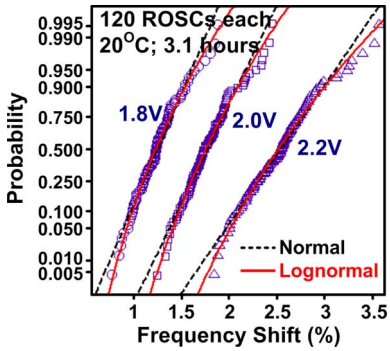


Fig. 14. While Rauch found that the distribution of ΔV_{th} values followed a Skellam distribution with tails that fall in between those of the normal and lognormal cases seen here [3], our Δf values fit well to the lognormal distribution.

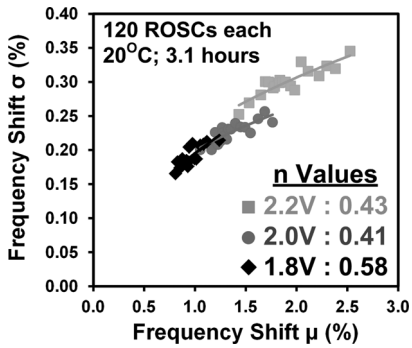


Fig. 15. Power laws with exponents ranging from 0.41 to 0.58 were fit to these $\sigma(\Delta f)$ versus $\mu(\Delta f)$ plots, matching the results found by Pae [2].

those of the normal and lognormal distributions [3], but Fischer later observed a larger shift in scaled SRAM devices [11]. Larger sample sizes and consistent measurement techniques are needed to clarify this discrepancy. Note that throughout this work, we still use the normal distribution to calculate the standard deviation, as Pae did [2]. Fig. 15 shows power laws with exponents of 0.41–0.58 fit to the $\sigma(\Delta f)$ versus $\mu(\Delta f)$ characteristics from our DC stress measurements. Pae also found exponents of 0.4-0.5 fit his ΔV_{th} and Δf results, and stated that this matches the prediction found by analytical equations after assuming a Poisson distribution for the number of charges created during stress. Kaczer derived a model that supports this power law dependence of 0.5 as well [4].

D. Temperature Dependence of DC Stress-Induced Degradation

Fig. 16 displays the degradation characteristics of the μ and σ of the frequency shifts at high temperatures. The power law exponents of these values increase at higher temperatures, and that of σ is just over 1/2 that of the μ characteristic. Varghese *et al.* stated that a linear dependence of n on temperature points to dispersive temperature dependence rather than Arrhenius activation, and that this phenomenon is simply an artifact of long measurement times [33]. They showed results indicating that the temperature dependence of n disappears with on-the-fly measurements. Other recent work by Liu *et al.* showed n increasing

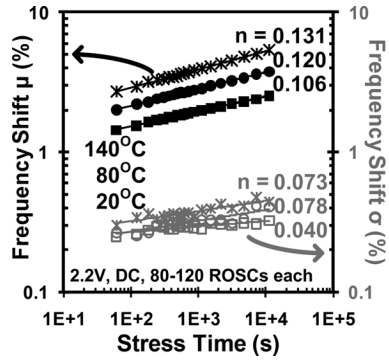


Fig. 16. Mean and standard deviation of the measured frequency shift at increasing temperatures.

with temperatures up to roughly 110°C, where the value saturated at 0.18, even when using fast and on-the-fly methods [34].

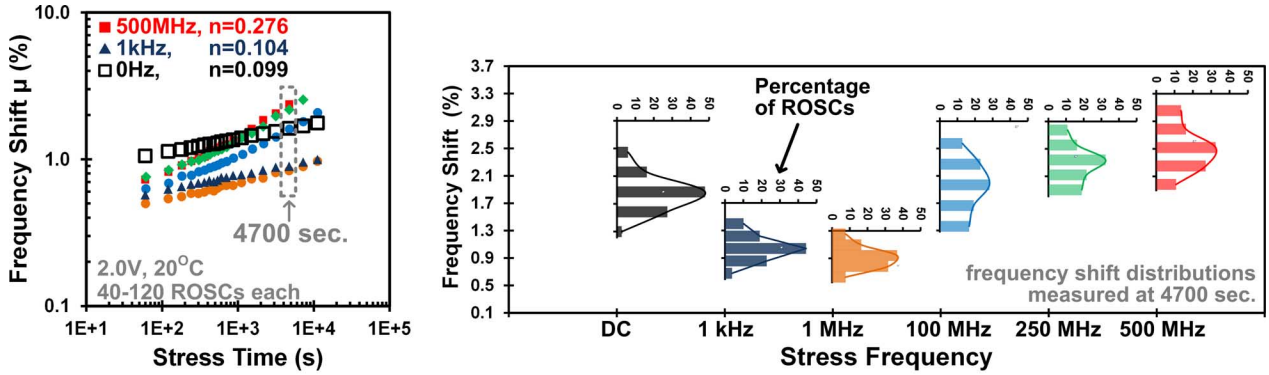
E. AC Stress and Stress/Recovery Characteristics

Fig. 17(a) shows a drop in total degradation of roughly 45% at low frequencies, compared with DC stress, due to the recovery that takes place during each half cycle for all PMOS. During DC stress half of the PMOS devices are stressed as mentioned earlier. On the other hand, under AC stress all PMOS devices experience equal amounts of stress and recovery, leaving them with only the damage that does not have time to recover along with any hot carrier-induced damage. If we estimate the impact of DC stress on *all* PMOS in the ROSC by multiplying that measured shift by two, we find a (1 kHz AC)/(DC) degradation ratio of $\sim$ 28% throughout the stress time. This value is just below the low range of 30–40% noted by Reisinger *et al.* [21].

As the frequency is raised, HCI plays a larger role in the aging due to the increased switching activity. This leads to a larger n, which is a signature of HCI [35]. At high voltages, we see that HCI eventually dominates the overall aging of the DUTs when the AC stress lines cross the DC characteristic. However, we have shown in previous work that this crossover point is highly dependent on voltage, and NBTI is dominant at lower stress voltages, closer to those used in real operation [16]. Note that this analysis is not equivalent to those found in pure AC BTI experiments, where no current flows through the channel, even during switching. Fernandez *et al.* found pure NBTI to be frequency independent up to 2 GHz [36].

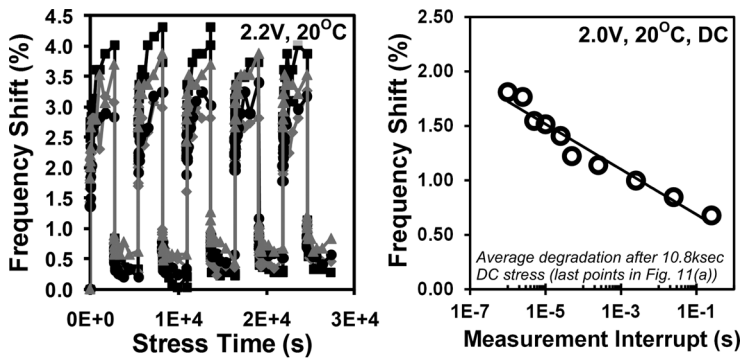
Histograms of these shifts at the 4700 s point are presented in Fig. 17(b). As in the DC case, no correlation was found between the fresh DUT frequency and the total frequency shift. The σ of the frequency shift again increases with stress, at a rate that increases with frequency. Although this curve only roughly follows a power law (the R-squared value is 0.85 at 500 MHz), the exponent values are around one half of that of the μ shift. Finally, the σ of the calculated frequency remained nearly flat on average, as explained for DC stress in Section IV-C.

Stress/Recovery curves taken from four ROSCs simultaneously are presented in Fig. 18(a). The bottom point of the recovery phases increases on average with each period as damage accumulates. The fast and significant recovery we observe after stress conditions are removed has been detected with other fast



(a) (b)

Fig. 17. (a) Mean AC stress results compared with DC. (b) Histograms of the frequency shift at the 4700 s point in part (a).



(a) (b)

Fig. 18. (a) Stress/Recovery curves taken from four ROSCs simultaneously. The bottom point of the recovery phases increases with each period on average as more long-term or permanent damage accumulates. (b) Logarithmic recovery characteristic constructed with data from different measurement times (i.e., stress interruptions) in Fig. 11 after 10,800 seconds.

measurement setups, and many state that it cannot be correctly described by the RD model [7], [10], [23], [26]. They state that one must use a fast hole trapping/detrapping model to find a theoretical explanation for these dynamics. Fig. 18(b) shows a logarithmic recovery characteristic constructed with data from different measurement times (i.e., stress interruptions) in Fig. 11 after 10,860 seconds of stress. Although this plot shows averaged data from multiple ROSCs and dies for each plotted point rather than traces of device characteristics throughout an extended recovery time, the logarithmic characteristic is common to both cases [5], [7].

V. Conclusions

We have implemented a measurement system that facilitates efficient statistical aging experiments involving BTI and HCI in ROSCs. Measurement results from a 65 nm test chip show that the differential Silicon Odometer beat frequency detection system can measure frequency shifts with an error of down to 0.07%, and stress interruptions of $\geq$ 1 μs. Statistical results show that fresh frequency and the AC or DC stress-induced frequency shift are uncorrelated, both the μ and σ of that shift increase with stress, and the ratio of this σ/μ decreases with stress time. These findings point to the utility of our proposed system for process characterization, and important trends in the aging of vanishingly small modern transistors.

References

[1] M. Agostinelli, S. Pae, W. Yang, C. Prasad, D. Kencke, S. Ramey, E. Snyder, S. Kashyap, and M. Jones, "Random charge effects for PMOS NBTI in ultra-small gate area devices," in *IEEE Int. Reliability Physics Symp.*, 2005, pp. 529–532.

[2] S. Pae, J. Maiz, C. Prasad, and B. Woolery, "Effect of BTI degradation on transistor variability in advanced semiconductor technologies," *IEEE Trans. Device Materials and Reliability*, vol. 8, no. 3, pp. 519–525, Sep. 2008.

[3] S. Rauch, "Review and reexamination of reliability effects related to NBTI-induced statistical variations," *IEEE Trans. Device Materials and Reliability*, vol. 7, no. 4, pp. 524–530, Dec. 2007.

[4] B. Kaczer, T. Grasser, P. Roussel, J. Franco, R. Degraeve, L. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," in *IEEE Int. Reliability Physics Symp.*, 2010, pp. 26–32.

[5] B. Kaczer, T. Grasser, J. Martin-Martinez, E. Simoen, M. Aoulaiche, P. Roussel, and G. Groeseneken, "NBTI from the perspective of defect states with widely distributed time scales," in *IEEE Int. Reliability Physics Symp.*, 2009, pp. 55–60.

[6] H. Reisinger, T. Grasser, and C. Schlünder, "A study of NBTI by the statistical analysis of the properties of individual defects in pMOSFETs," in *IEEE Int. Integrated Reliability Workshop*, 2009, pp. 30–35.

[7] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, "Recent advances in understanding the bias temperature instability," in *IEEE Int. Electron Devices Meeting (IEDM)*, 2010, pp. 4.4.1–4.4.4.

[8] S. Rauch, "The statistics of NBTI-induced VT and β mismatch shifts in pMOSFETs," *IEEE Trans. Device and Materials Reliability*, vol. 2, no. 4, pp. 89–93, Dec. 2002.

[9] G. La Rosa, W. Ng, and S. Rauch, "Impact of NBTI induced statistical variation to SRAM cell stability," in *IEEE Int. Reliability Physics Symp.*, 2006, pp. 274–282.
[10] V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes, and L. Camus, "NBTI degradation: From transistor to SRAM arrays," in *IEEE Int. Reliability Physics Symp.*, 2008, pp. 289–300.
[11] T. Fischer, E. Amirante, K. Hofmann, M. Ostermayr, P. Huber, and D. Schmitt-Landsiedel, "A 65 nm test structure for the analysis of NBTI induced statistical variation in SRAM transistors," in *IEEE European Solid-State Device Research Conf. (ESSDERC)*, 2008, pp. 51–54.
[12] K. Kang, S. Park, K. Roy, and M. Alam, "Estimation of statistical variation in temporal NBTI degradation and its impact on lifetime circuit performance," in *IEEE/ACM Int. Conf. Computer-Aided Design*, 2007, pp. 730–734.
[13] C. Tu, S. Chen, A. Chuang, H. Huang, Z. Jhou, C. Chang, S. Chou, and J. Ko, "Transistor variability after CHC and NBTI stress in 90 nm pMOSFET technology," *IEEE Electronics Lett.*, vol. 45, no. 16, pp. 854–856, Jul. 2009.
[14] V. Huard, "Two independent components modeling for negative bias temperature instability," in *IEEE Int. Reliability Physics Symp.*, 2010, pp. 2A.4.1–2A.4.10.
[15] 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.
[16] J. Keane, X. Wang, D. Persaud, and C. H. Kim, "An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB," *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 817–829, Apr. 2010.
[17] J. Keane, W. Zhang, and C. H. Kim, "An on-chip monitor for statistically significant circuit aging characterization," in *IEEE Int. Electron Devices Meeting*, 2010, pp. 4.2.1–4.2.4.
[18] K. Hofmann, H. Reisinger, K. Ermisch, C. Schlunder, W. Gusting, T. Pompl, G. Georgakos, K. v. Arnim, J. Hatsch, T. Kodytek, T. Baumann, and C. Pacha, "Highly accurate product-level aging monitoring in 40 nm CMOS," in *IEEE Symp. VLSI Technology*, 2010, pp. 27–28.
[19] E. Y. Wu, E. Nowak, A. Vayshenker, W. L. Lai, and D. L. Harmon, "CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics," *IBM J. Research and Development*, pp. 287–298, Mar./May 2002.
[20] W. Wang, V. Reddy, B. Yang, V. Balakrishnan, S. Krishnan, and Y. Cao, "Statistical prediction of circuit aging under process variations," in *IEEE Custom Integrated Circuits Conf. (CICC)*, 2008, pp. 13–16.
[21] H. Reisinger, T. Grasser, K. Hofmann, W. Gustin, and C. Schlunder, "The impact of recovery on BTI reliability assessments," in *IEEE Int. Integrated Reliability Workshop*, 2010, pp. 12–16.
[22] T. Grasser, P. Wagner, P. Hehenberger, W. Goes, and B. Kaczer, "A rigorous study of measurement techniques for negative bias temperature instability," *IEEE Trans. Device and Materials Reliability*, vol. 8, no. 3, pp. 526–535, Sep. 2008.
[23] H. Reisinger, R.-P. Vollertsen, P.-J. Wagner, T. Huttner, A. Martin, S. Aresu, W. Gustin, T. Grasser, and C. Schlunder, "A study of NBTI and short-term threshold hysteresis of thin nitrided and thick non-nitrided oxides," *IEEE Trans. Device Materials and Reliability*, vol. 9, no. 2, pp. 106–114, Jun. 2009.
[24] Z. Ji, J. Zhang, M. Chang, B. Kaczer, and G. Groeseneken, "An analysis of the NBTI-induced threshold voltage shift evaluated by different techniques," *IEEE Trans. Electron Devices*, vol. 56, no. 5, pp. 1086–1093, May 2009.
[25] M.-F. Li, D. Huang, C. Shen, T. Yang, W. Liu, and Z. Liu, "Understand NBTI mechanism by developing novel measurement techniques," *IEEE Trans. Device and Materials Reliability*, vol. 8, no. 1, pp. 62–71, Mar. 2008.
[26] C. Shen, M.-F. Li, C. Foo, T. Yang, D. Huang, A. Yap, G. Samudra, and Y.-C. Yeo, "Characterization and physical origin of fast V_{th} transient in NBTI of pMOSFETs with SiON dielectrics," in *IEEE Int. Electron Devices Meeting*, Dec. 2006, pp. 1–4.
[27] H. Reisinger, T. Grasser, W. Gustin, and C. Schlunder, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress," in *IEEE Int. Reliability Physics Symp.*, 2010, pp. 2A.1.1–2A.1.9.
[28] D. Ielmini, M. Manigrasso, F. Gattel, and M. G. Valentini, "A new NBTI model based on hole trapping and structural relaxation in MOS dielectrics," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1943–1952, Sep. 2009.
[29] Y. Wang, "Effects of interface states and positive charges on NBTI in silicon-oxynitride p-MOSFETs," *IEEE Trans. Device and Materials Reliability*, vol. 8, no. 1, pp. 14–21, Mar. 2008.
[30] S. Mahapatra and M. Alam, "Defect generation in p-MOSFETs under negative-bias stress: An experimental perspective," *IEEE. Trans. Device and Materials Reliability*, vol. 8, no. 1, pp. 35–46, Mar. 2008.
[31] S. Mahapatra, V. Maheta, A. Islam, and M. Alam, "Isolation of NBTI stress generated interface trap and hole-trapping components in PNO p-MOSFETs," *IEEE. Trans. Electron Devices*, vol. 56, no. 2, pp. 236–242, Feb. 2009.
[32] E. Saneyoshi, K. Nose, and M. Mizuno, "A precise-tracking NBTI-degradation monitor independent of NBTI recovery effect," in *IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2010, pp. 192–193.
[33] D. Varghese, D. Saha, S. Mahapatra, K. Ahmed, F. Nouri, and M. Alam, "On the dispersive versus Arrhenius temperature activation of NBTI time evolution in plasma nitrided gate oxides: Measurements, theory, and implications," in *IEEE Int. Electron Devices Meeting (IEDM)*, 2005, pp. 684–687.
[34] W. Liu, D. Huang, Q. Sun, C. Liao, L. Zhang, Z. Gan, W. Wong, and M.-F. Li, "Studies of NBTI in pMOSFETs with thermal and plasma nitrided SiON gate oxides by OFIT and FPM methods," in *IEEE Int. Reliability Physics Symp. (IRSP)*, 2009, pp. 964–968.
[35] H. Kufluoglu, "MOSFET degradation due to negative bias temperature instability (NBTI) and hot carrier injection (HCI) and its implications for reliability aware VLSI design," Ph.D. dissertation, Purdue Univ., West Lafayette, IN, 2007.
[36] R. Fernández, 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 *IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2006, pp. 337–340.

John Keane (S'06–M'11) received the B.S. degree in computer engineering from the University of Notre Dame, Notre Dame, IN, in 2003, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2010.

He is currently with Advanced Design in the Portland Technology Development group of Intel Corporation, Hillsboro, OR, where he is working on memory test circuits. He has coauthored over 20 peer-reviewed publications and is a co-inventor on five US patents.

Dr. Keane won 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, and served on the technical program committee for the 2011 International Symposium on Low Power Electronics and Design (ISLPED).

Wei Zhang (S'11) received the B.S. degree in electrical engineering from Tsinghua University, China. He is currently a Ph.D. candidate at the University of Minnesota, Minneapolis, where he has been involved in five chip designs since joining Professor Chris Kim's research group in 2007.

His work is focused on low power circuit design techniques as well as SRAM and embedded DRAM design. He is also involved in a cooperative project on organic electronics. He spent five months at Broadcom Corporation as an intern engineer where he conducted research on memory-related designs in the summers of 2009 and 2010. He is the author or coauthor of seven journal and conference papers.

Chris H. Kim (M'04–SM'10) received the B.S. and M.S. degrees from Seoul National University, Seoul, Korea, and the Ph.D. degree 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 at the University of Minnesota, Minneapolis, in 2004 where he is currently an Associate Professor.

Prof. Kim is the recipient of an NSF CAREER Award, a McKnight Foundation Land-Grant Professorship, a 3M Non-Tenured Faculty Award, DAC/ISSCC Student Design Contest Awards, IBM Faculty Partnership Awards, an IEEE Circuits and Systems Society Outstanding Young Author Award, ISLPED Low Power Design Contest Awards, an Intel Ph.D. Fellowship, and the Magoon's Award for Excellence in Teaching. He is an author/coauthor of more than 90 journal and conference papers and has served as a technical program committee member for numerous circuit design conferences. He was the technical program committee chair for the 2010 International Symposium on Low Power Electronics and Design (ISLPED) and the guest editor for a special issue of the IEEE *Design and Test Magazine*. His research interests include digital, mixed-signal, and memory circuit design in silicon and non-silicon (organic and magnetic) technologies.