

An On-Chip Monitor for Statistically Significant Circuit Aging Characterization

John Keane^{1,2} Wei Zhang¹ Chris H. Kim¹

1. Department of ECE, University of Minnesota, 200 Union Street Southeast, Minneapolis, MN 55455, USA
 2. Intel Corporation, 2501 Northwest 229th Avenue, Hillsboro, OR 97124, USA (Email: john.keane@intel.com)

Abstract

We present an array-based system to accurately study variations in ROSC aging. Microsecond measurements for minimal BTI recovery, and 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 65nm test chip show that fresh frequency and Δf are uncorrelated, both the μ and σ of Δf increase with stress, and $\sigma(\Delta f)/\mu(\Delta f)$ decreases with stress time.

Introduction

Transistor aging is the product of a finite number of trapped charges or broken bonds in exceedingly minute modern devices, so it is no longer sufficient to rely on a small set of stress tests to predict the behavior of billions of devices over the lifetime of a circuit. Statistical fluctuations in the number and positions of defects contributing to transistor degradation in scaled devices lead to a distribution of effective device “ages” at any given time. 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 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 induced 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 induce in the measured parameter of interest (e.g., V_{TH}), is also randomly distributed. That step size is dependent on each charge’s position relative to the others [3]. Some authors have also claimed that the widely distributed time constants of the defects created during BTI stress contribute to the variation [4]. For these reasons, variations in transistor aging have received increasing attention as CMOS technology has been pushed into the deep sub-micron regime.

In this paper, we present the first array-based system to accurately study variations in aging involving the latter two mechanisms in stressed ring oscillators (ROSC). The distribution of frequencies is monitored by a set of three Silicon Odometer beat frequency detection systems working in parallel [5-6]. Unwanted BTI recovery during stress interruptions is avoided with measurements of $\geq 1\mu s$. Frequency shift (Δf) measurement resolution ranging down to the error floor of 0.07% is achieved in combination with those quick measurements, which is not possible with standard off-chip test equipment. In addition to these timing and measurement resolution benefits, on-chip stress testing can provide area savings (since stressed devices share I/O circuits

and pads), as well as important insight into the degradation of realistic circuits rather than individual devices. Results from a 65nm test chip show that fresh frequency and Δf are uncorrelated, both the average (μ) and standard deviation (σ) of Δf increase with stress, and $\sigma(\Delta f)/\mu(\Delta f)$ decreases with stress time.

Statistical Circuit Aging Monitor

A block diagram of the proposed design is shown in Fig. 1. The references are identical to the ROSCs within the array (Fig. 2), and are left unstressed to maintain fresh reference points in the Odometers’ differential measurement setup. Ten inverters in each ROSC are 1.2V thin oxide logic devices. These stages will age during stress experiments, while the rest of the stress and timing control logic is composed of 2.5V thick oxide I/O transistors, so it is not significantly impacted. Measurements are taken during circuit calibration to calculate the percentage of the fresh full loop delay accounted for by the logic devices under test (DUT). Later, the total frequency shift of each stressed full loop measured by the Odometers is divided by the percentage of the fresh delay taken by the

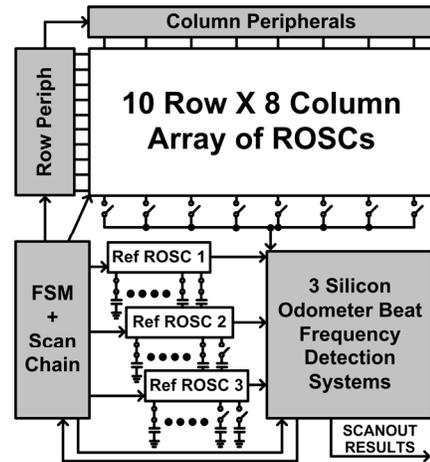


Fig. 1. System diagram. Reference ROSCs have 15 trimming capacitors controlled by the scan chain. Stressed ROSCs are taken out of stress individually for measurements using the FSM and Peripheral circuits.

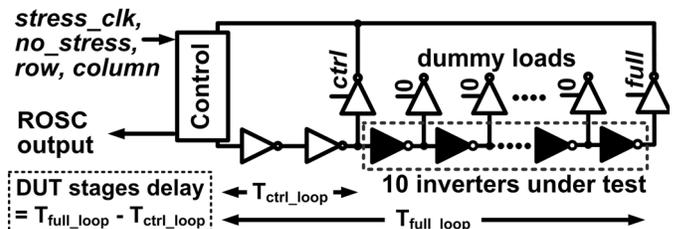


Fig. 2. Basic structure of the ROSC cells. Stages highlighted in black are composed of thin oxide logic devices. All others are 2.5V I/O devices.

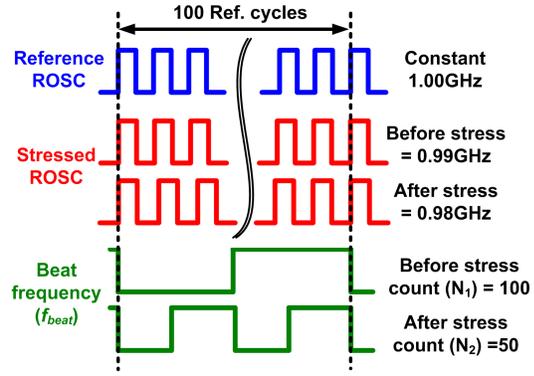
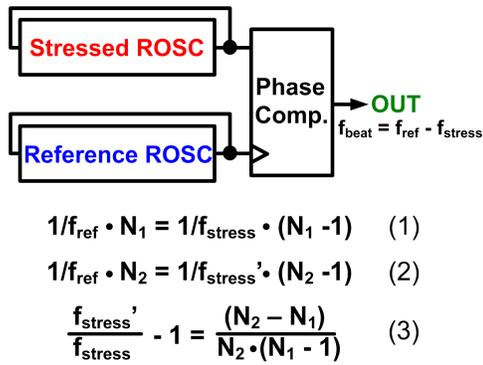


Fig. 3. Concept behind the beat frequency detection system. Equation 1 is valid before stress, and the (N_1-1) term arises from the fact that the slower stressed ROSC will take one less period to cycle back to the same point in the reference ROSC period while both are oscillating. After stress, only the count and f_{stress} change (Equation 2). Using these two equations, we can calculate the stressed ROSC frequency shift, as shown in Equation 3. The measurement time is limited only by the reference ROSC period and the number of cycles that must be counted [6].

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. 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, its supply is set to the standard digital level of 1.2V, the loop is closed, and its frequency shift is measured by the three Odometer systems.

The Silicon Odometer measures frequency changes in the stressed ROSCs with the concept illustrated in Fig. 3 [5-6]. 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: $f_{\text{beat}} = f_{\text{ref}} - f_{\text{stress}}$. 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 the stressed ROSC frequency with straight-forward algebraic equations.

Silicon Odometers provide high-resolution Δf measurements when the frequencies of the ROSC under test and reference are close because this results in a low beat frequency, and hence, a high output count (see Fig. 3). This is ensured with trimming capacitors. However, in the present circuit where many ROSCs are stressed in parallel and selected one-by-one for measurements, controlling the trimming bits in each stressed oscillator would be time and area consuming. Therefore, we instead hardwired nine of

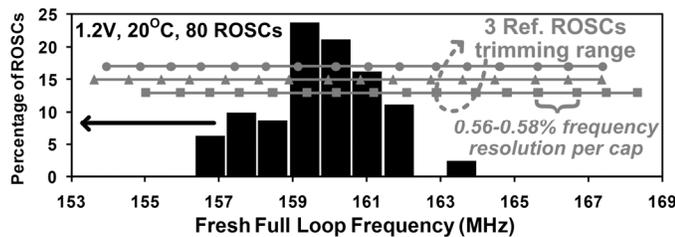


Fig. 4. Fresh full loop frequency distribution from 80 cell array, with Reference ROSC trimming range. The reference ROSCs are trimmed to positions within the distribution such that we maximize the resolution of the degradation characteristics gathered from each full experiment.

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 output counts corresponding to the beat frequency for each ROSC measurement (Figs. 1 and 4). During post-processing, the highest-resolution degradation characteristic is selected from that set for each ROSC that was stressed.

Test Chip Aging Measurements

A test circuit was implemented in 65nm CMOS for concept verification. We first performed 0V, no-stress experiments, so ideally there should be no Δf . Single-ended measurements recorded directly by a 100MHz, 1.25GS/s oscilloscope, after frequencies were divided down by 1024 on chip, had a worst case error of 0.17% due slight variations in operating conditions. On-chip Odometer measurements had a smaller worst case error of 0.07%, in part because this differential system eliminates the effects of common mode variations. Finally, note that the automated oscilloscope readings required over 500ms, while the Odometer measurements take down to 1 μ s in the current design (see our previous publication for more details on the time required for odometer measurements [6]). This combination of resolution and fast measurements is critical when measuring BTI stress,

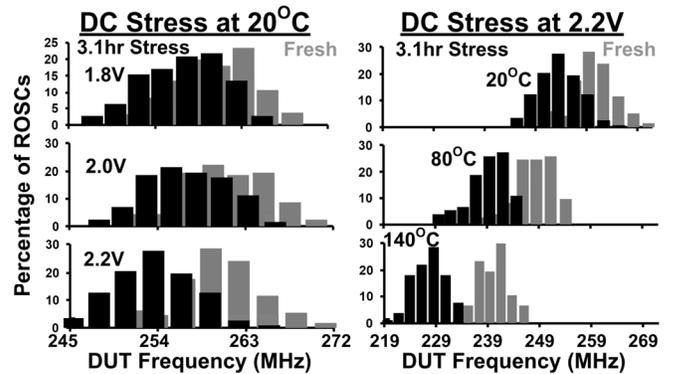


Fig. 5. Shift in frequency distributions after 3.1 hour stress. Each 20°C distribution was gathered from 120 ROSCs, while those at higher temperatures came from 80 due to a limited number of dies.

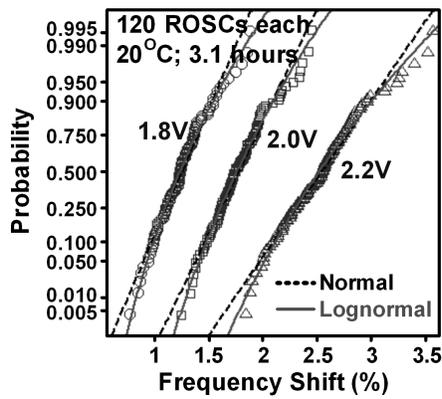


Fig. 6. Measured Δf values fit well to the lognormal distribution. The high tails are not overestimated by the lognormal here unlike the threshold shifts measured by Rauch [3].

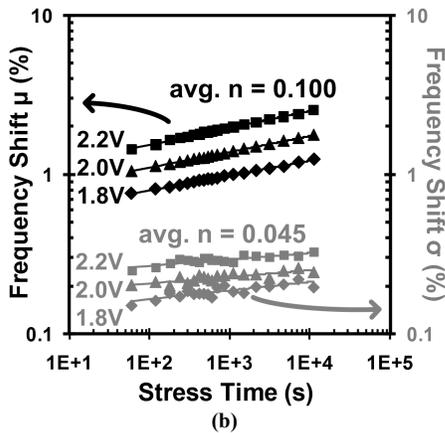
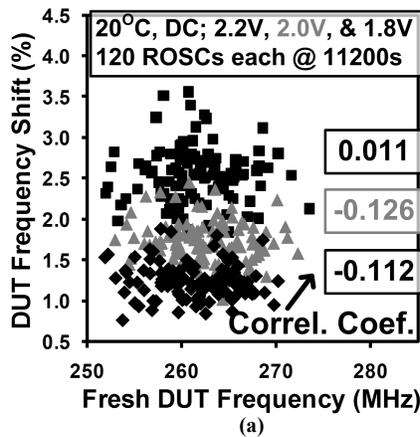


Fig. 7. (a) No significant correlation was found between the stress-induced Δf and the fresh frequencies. Pae *et al.* found that the initial spread in the threshold voltage is not correlated to the shift caused by aging [2]. (b) The mean and standard deviation of Δf both increase with stress time.

where we must avoid recovery when stress conditions are temporarily removed for readings.

PDFs of fresh DUT frequencies are shown in Fig. 5 with the resulting distributions after 3.1 hours of DC stress. The primary degradation characteristic at work in these experiments was NBTI, since PBTI is not significant with SiO_2 dielectrics [2], and there was no switching during stress. In Fig. 6, we see that the measured frequency shift

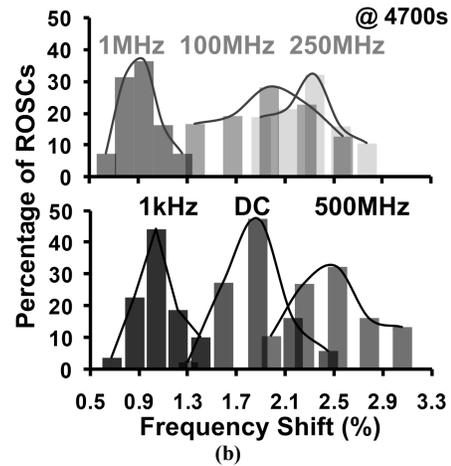
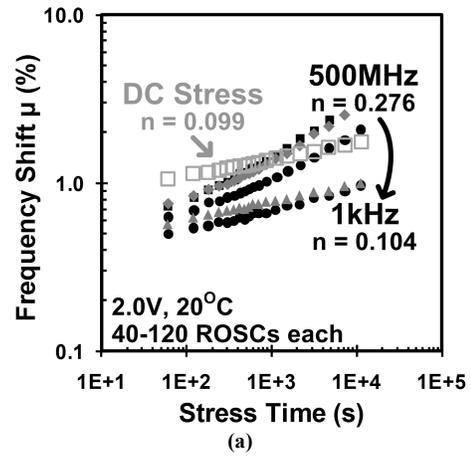


Fig. 8. (a) Mean AC stress results compared with the DC case. We see a drop in degradation of roughly half at low frequencies compared with DC stress due to BTI recovery. As frequency is raised, HCI plays a larger role in the aging. This leads to a larger n , which is a signature of HCI [6]. (b) PDFs of the frequency shift at the 4700s point in part (a).

distributions fit well to a lognormal distribution. Rauch found that the lognormal overestimated the high tails of his measured V_{TH} shifts, and that his data fit well to a Skellam with tails in between those of the normal and lognormal distributions [3]. Larger sample sizes may be needed to clarify this discrepancy. Note that throughout this work, we still use the normal distribution to calculate the standard deviation, as Pae *et al.* did [2].

In Fig. 7(a) we see that there was no significant 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 [7], and that the initial spread in the V_{TH} is not correlated to that caused by aging [2]. Fig. 7(b) shows the μ frequency shifts and the σ of the shifts vs. stress time. The σ increases with stress, roughly following a power law with an exponent (n) of just under 1/2 that of the μ shift. Therefore, this $\sigma(\Delta f)/\mu(\Delta f)$ ratio decreases with stress time.

Fig 8(a) illustrates a drop in total degradation of $\sim 1/2$ at low frequencies, compared with DC stress, due to the NBTI recovery that takes place during each half cycle for all PMOS. As the frequency is raised, HCI plays a larger role in the aging

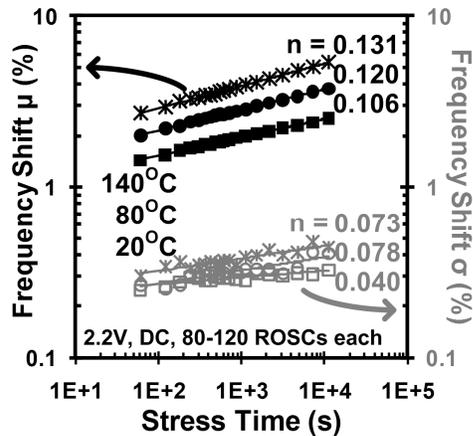


Fig. 9. High temperature results. The increase in the power law exponent points to dispersive temperature dependence rather than Arrhenius activation [8].

due to the increased switching activity. This leads to a larger power law exponent, which is a signature of HCI [6]. At high voltages, 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 found in real operation [6]. PDFs of these shifts at the 4700s point are presented in Fig. 8(b).

Fig. 9 shows 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 surpasses 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 [8]. 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 with temperatures up to roughly 110°C, where the value saturated at 0.18, even when using fast and on-the-fly methods [9]. Like many issues surrounding BTI degradation, this matter is unsettled, but our results from a large sample set support the dispersive transport model.

Fig 10 quantifies the impact of measurement times on BTI measurements. Long interruptions 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. Several previous publications have clearly demonstrated this phenomenon [10]-[13]. As times get into the millisecond range, the average exponent approaches 0.167, which has commonly been cited as the measured and theoretically correct value. It fits the RD model theory if H_2 is assumed to be the hydrogen species diffusing away from the silicon/dielectric interface [13]. However, our results, along with those from several other fast measurement techniques, show that this model seems to be incomplete or

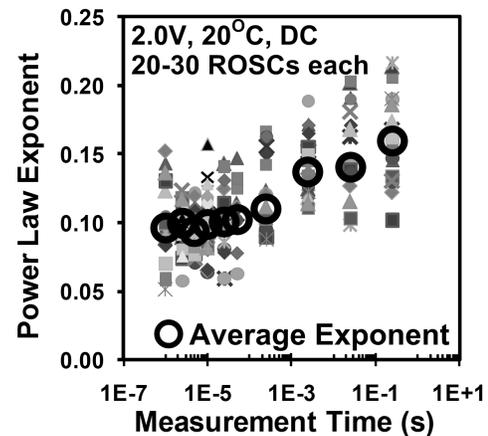


Fig. 10. Power law exponent increase with stress interruption time. As times get into the millisecond range, the mean exponent approaches 0.167, which has commonly been cited as the correct value [13].

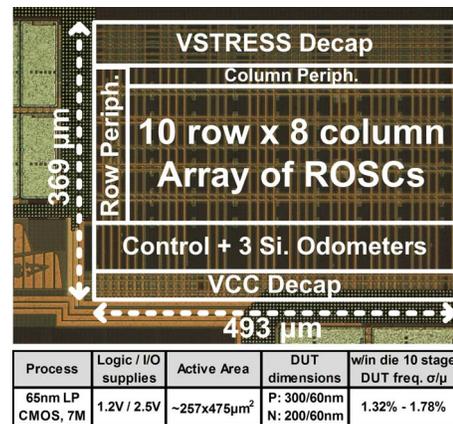


Fig. 11. Die microphotograph and test chip characteristics. Measurements were automated with LabVIEW™ software through a National Instruments data acquisition board.

incorrect [10], [13], [14]. Measurement times of tens of μs or less are required to observe the power law exponent of ~ 0.1 .

Acknowledgements

This work was supported in part by the SRC under award 2008-HJ-1805. The authors would also like to thank Intel, Broadcom, and the IBM PhD Fellowship program for their financial support.

References

- [1] M. Agostinelli, *et al.*, IRPS, pp. 529-532, 2005.
- [2] S. Pae, *et al.*, TDMR, vol. 8, no. 3, pp. 519-525, Sept. 2008.
- [3] S. Rauch, *et al.*, TDMR, vol. 7, no. 4, pp. 524-530, Dec. 2007.
- [4] B. Kaczer, *et al.*, IRPS, pp. 55-60, 2009.
- [5] T. H. Kim, *et al.*, JSSC, vol. 43, no. 4, pp. 874-880, Apr. 2008.
- [6] J. Keane, *et al.*, JSSC, vol. 45, no. 4, pp. 817-829, Apr. 2010.
- [7] S. Rauch, TDMR, vol. 2, no. 4, pp. 89-93, December 2002.
- [8] D. Varghese, *et al.*, IEDM, pp. 684-687, 2005.
- [9] W. Liu, *et al.*, IRPS, pp. 964-968, 2009.
- [10] T. Grasser, *et al.*, TDMR, vol. 8, no. 3, pp. 526-535, September 2008.
- [11] H. Reisinger, *et al.*, TDMR, vol. 9, no. 2, pp. 106-114, June 2009.
- [12] Z. Ji, *et al.*, TED, vol. 56, no. 5, pp. 1086-1093, May 2009.
- [13] M.-F. Li, *et al.*, TDMR, vol. 8, no. 1, pp. 62-71, March 2008.
- [14] C. Shen, *et al.*, IEDM, pp. 1-4, December 2006.