Towards Wearout-aware and Accelerated Self-healing Digital Systems

I. AUTHOR INFORMATION

- **Track**: 4. Power and Reliability Analysis and Optimization
- **Author Name**: Xinfei Guo
- **Institution**: University of Virginia, Dept. of ECE
- **Author Email**: xg2dt@virginia.edu
- **Advisor**: Prof. Mircea R. Stan
- **Advisor Email**: mircea@virginia.edu
- **Graduation Date**: December 2016 (Expected)
- **This work has not been presented at ASP-DAC Ph.D. Forum nor DATE Ph.D. Forum.**

II. DISSERTATION ABSTRACT

**Background**: Wearout (Aging) has been one of the most significant reliability concerns in CMOS circuit as the technology scaling continues and technology node advances [1]. It is a long term process that is caused by severe interrelated physical mechanisms that conspire to worsen system metrics by slowing the circuit down and increasing the power. Even though emerging technologies, such as FinFET and Gate-all-around (GAA) silicon nanowire (SiNW) FET, with excellent electrostatic properties and comparative ease of fabrication, appear to be good candidates for further extending the technology to the nanoscale regime, wearout is still a big issue and it becomes even worse in these devices due to self-heating effects and continuous oxide thickness scaling requirements [2, 3].

**Motivation**: Wearout consists of both reversible and irreversible (permanent) parts which accumulate at different rates under stress, when the stress is removed, there is some level of recovery, but usually at much lower rate than the wearout. In the front-end of line (FEOL), Bias Temperature Instability (BTI) is one of the most prominent reversible wearout effects that shift transistor’s threshold voltage and further slow the circuit down. Similar shifts happen in the back-end of line (BEOL) for interconnect, with electromigration (EM) being the main wearout phenomenon that shortens metal wire lifetime, and become severe due to power delivery challenges for large ICs. In the past decade, lots of work has been done in the community to deal with wearout issues from system level down to device level. There are two ways in general, one way is to relax the design margin by designing guard banding and/or design for the worst case [4], and the second way is being adaptive to the wearout induced variations by tracking, monitoring and compensating dynamically [5], thus being able to design for the average case. The issue of both methods is that with scaling, the worst case becomes even worse and the distribution becomes skewed, thus the potential advantages of adaptation are also reduced, which means the system might function correctly with adaptation, but will still become sluggish and burn too much power, especially during the early lifetime. A better solution would be to somehow reduce the actual wearout induced variations. One way is to reduce the stress, thus to alleviate wearout during run time [6], it is obvious that this method will introduce performance overhead. In this work, we propose and demonstrate a more fundamental solution inspired by circadian rhythms called accelerated self-healing that actually repairs wearout periodically by taking advantage of recovery property of several wearout effects, such as BTI for transistors, EM for metal wires, etc. The main objective of this work is to improve the lifetime as well as the PPA metrics of future electronic systems by the use of extensive and accelerated recovery methods, such as high temperatures, negative voltages and currents, UV exposure, periodical rejuvenation.

**Goals**: The goals of our research are the following: 1. Understand and model the recovery behaviors of several wearout mechanisms, especially BTI and EM, with and without several accelerated self-healing techniques; 2. Develop and evaluate the on-chip solutions for the proposed accelerated self-healing techniques; 3. Explore the circadian rhythm behaviors with proactive periodical accelerated rejuvenation, and develop a cross-layer optimization infrastructure that allows the design, analysis and optimization of systems across hierarchy.

**Accelerated Self-healing**: Here we rephrase the notion of sleep for electronic systems since until now, sleep for systems is widely accepted as a period of inactivity or idleness, quite different from biological organisms which, during sleep, go through several essential active processes for the recovery of their full capabilities for the next day. In this work, we borrow the circadian rhythms inspired definitions for sleep, and demonstrate that sleep could actually be used as an active recovery period from wearout. During sleep, some of the effects (e.g. BTI and EM) can be reversed heavily by several techniques, such as applying high temperature and negative supply voltages, reversing the current direction flowing through the metal wire, using UV light exposure, thus leading to effective accelerated self-healing. The sleep behaviors are studied on commercial FPGA chips fabricated in 40nm node, and the test structures are LUT-based ring oscillators that capture the frequency change during active and sleep periods. As shown in the following figure, we demonstrate a test case where 72.4% of the wearout is recovered within only 1/4 of the stress time [C6]. This improvement will be directly translated to improvements for some metrics, such as relaxed design margin.

---

**Image**

![Fig. 1 Illustration of wearout vs. accelerated self-healing](image)

---

**This work is funded partially by NSF No. CCF-1255907, and by SRC through Global Research Collaboration (GRC) under task ID. 2410.001.**
Circadian Rhythms: To explore the circadian rhythm behavior, the comprehensive tests on the same FPGA chips are conducted. The testing results show that after an initial stress phase the circuits can get into an (almost) steady state where alternating phases of stress and accelerated recovery can compensate for each other, and after each recovery phase the chip can indeed start (almost) fresh. But there is a saturation of this effect once the recovery period exceeds a threshold. As an example we demonstrate a case where the scheduled active/accelerated recovery ratio is 4 to 1 which leads to an average performance improvement of 1.4% over a short period of only two days compared to the case where no recovery techniques are applied; this percentage would become proportionally higher for longer periods of time [C1]. Under such scenarios, like biological systems after “a good night’s sleep,” the system can start (almost) fresh after each recovery period, except for the small percentage of true irreversible wearout that cannot be recovered. A wearout-adaptation strategy now only needs to track rapid (reversible) wearout over a short period of time (e.g. 1 day instead of several years). A novel type of such sensors that can track both wearout and accelerated recovery is also proposed in this work [C3].

On-chip Implementations: To enable the on-chip implementation of the proposed accelerated self-healing techniques, we proposed several potential solutions. To generate on-chip heat, one method is to build some reconfigurable switching elements that could be used as self-heating blocks [9]. Another is that in multicore systems, some cores are active while others asleep for saving energy or for abiding by TDP limitations due to “Dark Silicon”. By taking advantage of the heat generated by active cores, the sleep cores can be fully rejuvenated. Negative voltages are widely used in assist circuit for SRAM and flash memories. A charge-pump based negative voltage generator is designed and embedded into a switch-cap DC-DC converter. The total area of the generator is only ~5% of the whole voltage regulation. We also propose that by applying a higher-than-nominal voltage at the gate of the header of power gating techniques during sleep mode, the header will be deeply rejuvenated due to the formation of negative voltage across the gate. A test chip in 28nm is being prepared for tapeout for validating these solutions, silicon measurements will be left as future work. In terms of UV light, there has been a gaining interest in heterogeneous 3D ICs [8] that enable stacking various technologies within a single device. We believe that this trend will lead to the embedded UV devices (e.g. [10]) in a System-on-Chip (SOC) system.

Cross-layer Optimization: While the physical phenomena that characterize wearout and accelerated recovery are at the device levels, their effects are apparent across the hierarchy, with timing or transient faults at circuit level, errors at the architecture level and failures at the system level. In addition, since the time before each scheduled deep rejuvenation is known in advance, there are good opportunities for cross-layer optimization that can take advantage of it. The only way to achieve optimal overall improvement is to develop circuit-level transient models [C2] that can be compatible with simulators (e.g. Spectre or SPICE), and physically aware parameterized high-level models that can be integrated with simulators like Gem5 [10] to evaluate the overall metric improvements – We are working on both of them. The success of the model will lead to the new design methodology, like design for accelerated recovery (DFAR) or Power- and Wearout-aware co-design methodology, and enable the optimized scheduling algorithms that trade off between lifetime and other metrics. The extension of the proposed methods to emerging technologies, such as FinFET and 3DIC, will also be investigated in the future.

III. DESCRIPTION OF SUPPORTING PAPER

The attached supporting paper [C2], Modeling and Experimental Demonstration of Accelerated Self-Healing Techniques, was presented at the 51st Design Automation Conference (DAC), where 174 out of 787 papers (22%) were accepted. In that paper, we proposed the notion of accelerated self-healing, modeled and demonstrated several rejuvenation techniques (higher temperatures, negative voltages, active/sleep ratio) by conducting comprehensive stress and recovery tests on commercial 40nm FPGAs. We bring stressed chips back to within 90% of their original margin by actively rejuvenating for only 1/4 of the stress time for all test cases.

IV. PUBLICATION LIST


REFERENCES