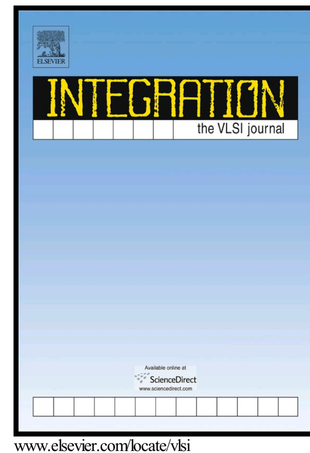
Author's Accepted Manuscript



www.elsevier.com/locate/vlsi

Implications of Accelerated Self-Healing as a Key Design Knob for Cross-Layer Resilience

Xinfei Guo, Mircea R. Stan

PII: S0167-9260(16)30084-0
DOI: http://dx.doi.org/10.1016/j.vlsi.2016.10.008
Reference: VLSI1254

To appear in: *Integration, the VLSI Journal*

Received date: 1 May 2016
Revised date: 14 September 2016
Accepted date: 12 October 2016

Cite this article as: Xinfei Guo and Mircea R. Stan, Implications of Accelerated Self-Healing as a Key Design Knob for Cross-Layer Resilience, *Integration, the VLSI Journal,* http://dx.doi.org/10.1016/j.vlsi.2016.10.008

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Implications of Accelerated Self-Healing as a Key Design Knob for Cross-Layer Resilience

Xinfei Guo*, Mircea R. Stan
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA 22904, USA
{xg2dt, mircea}@virginia.edu

***Abstract*—In this paper we propose a cross-layer accelerated self-healing (CLASH) system which "repairs" its wearout issues in a physical sense through accelerated and active recovery, by which wearout can be reversed while actively applying several accelerated self-healing techniques, such as high temperature and negative voltages. Different from previous solutions of coping with wearout issues (e.g. BTI) by "tolerating", "slowing down" or "compensating", which still leave the irreversible (permanent) wearout component unchecked, the proposed solution is able to fully avoid the irreversible wearout through periodic rejuvenation, and this is inspired by the explored frequency dependent behaviors of wearout and (accelerated and active) recovery based on measurements on FPGAs. We demonstrate a case where the chip can always be brought back to the fresh status by employing a pattern of 31-hour regular operation (under room temperature and nominal voltage) followed by a 1-hour accelerated self-healing (under high temperature and negative voltage). The proposed system integrates the notions of accelerated self-healing across multiple layers of the system stack. At the circuit level, a negative voltage generator and heating elements are designed and implemented; at the architecture level, the core can be allocated in a way such that the dark silicon or redundant resources can be healed by active elements; at the system level, right balance of stress and accelerated/active recovery can be employed by the system scheduler to fully mitigate the wearout; various wearout sensors act as the media between different layers. Overall, these techniques work together to guarantee that the whole system performs for more of the time at higher levels of performance and power efficiency by fully taking advantage of the extra opportunities enabled by the accelerated self-healing.**

***Index Terms*—Wearout, BTI, Accelerated self-healing, Frequency dependence, Cross-layer.**

I. INTRODUCTION

The never-ending demand for higher performance and lower power consumption pushes the aggressive technology scaling and the appearance of emerging devices, while further downscaling leads to major challenges, among which wearout (or aging) has become a huge reliability threat. Bias Temperature Instability (BTI) has been accepted as one of the most dominant wearout factors causing lifetime reliability problems in the front-end of line (FEOL) by worsening metrics across the digital system hierarchy [1]–[4], with performance degradation or intrinsic faults at the circuit level [5], errors at the architecture level [3] and failures at the system level [6]. Thus, dealing with wearout issues (such as BTI) needs to cross layers, where various techniques are necessary to be implemented - from device level up to the application level - to work together to achieve the optimal lifetime and acceptable wearout levels with a low cost [2], [4], [7], [8].

In general, these cross-layer techniques can be divided into two categories. The first one is during the design phase, when the worst case wearout levels are estimated (e.g. based on application behaviors) and design margin is reserved by adding guardband at the circuit level (e.g. oversizing), it has been shown in [2], [9], [10] that the margin can be $> 20\%$ for a 10-year lifetime constraint, the added margins usually lead to large timing slacks and therefore wasteful power consumption especially during the initial lifetime. A second solution would be adaptive techniques during run time, where wearout induced variations are tracked and monitored by sensors, then various actuators are employed for compensating or adapting to the variations dynamically [11], [12], so the system is able to be designed for the average case. Such actuators are not unlike those for PVT variations, for example, they can be DVFS and/or body bias at the circuit level [13], cache sizing at the architecture level [11], task allocation at the system level or loop perforation at the application level [3], [11]. While due to that wearout is time dependent by nature, and will get worse and worse fundamentally as the system runs even though some reversible (recoverable) wearout might not fully accumulate until the end of lifetime, thus the adaptation might be able to guarantee that the system function correctly, but it either runs sluggishly or burns too much power. In addition, wearout sensors (the expected number for a future SoC can reach as many as hundreds [11]) need to be ON for tracking over the entire lifetime, and this will add unacceptable tracking power overhead.

A better solution would be to somehow "repair" wearout issues by reducing the actual variations. Since many dominant wearout mechanisms, such as BTI, are voltage dependent, so one way is to scale down the voltage stress, thus to alleviate wearout during the run time [14], but this solution will introduce huge performance overhead. An alternative way is to take advantage of the recovery property of BTI by generating more idle time for *passive recovery* (system unstressed when not in use) [15]. While passive recovery is usually slow

*Corresponding author.
Address: High-performance Low-power (HPLP) Lab, Rice Hall 330, 85 Engineer's Way, Charlottesville, VA 22904, USA
Email address: xg2dt@virginia.edu (X. Guo),
mircea@virginia.edu (M. R. Stan)
Tel: +1-434-227-7800

Accepted manuscript

Download English Version:

https://daneshyari.com/en/article/4970743

Download Persian Version:

https://daneshyari.com/article/4970743

Daneshyari.com