

Overview of Health Monitoring Techniques for Reliability

Abhijit Deb, Bart Vermeulen, Luc van Dijk

NXP Semiconductors, Business Unit Automotive, The Netherlands

Email: { abhijit.deb | bart.vermeulen | luc.van.dijk } @nxp.com

Abstract—Many semiconductor circuit wearout monitoring techniques have been developed for different applications. This paper presents an overview of the prominent wearout monitoring techniques. We systematically categorize the semiconductor wearout monitoring techniques and qualitatively assess the different categories for the health monitoring application.

Keywords—ISO26262; reliability; NBTI; PBTI; HCI; wearout; aging; health monitoring.

I. INTRODUCTION

Semiconductor components are required to function reliably in safety critical applications in automotive and aeronautics domains. The harsh operating conditions and migration to newer technologies accelerate semiconductor wearout in the field. Thus aging poses a reliability risk. This problem is typically solved by allocating a wide enough timing margin to handle the worst-case conditions, which may however only occur rarely in practice. An alternative to avoid the overhead caused by these wider margins can be realized by novel run-time health monitoring of semiconductor circuits. The automotive safety standard, ISO26262, calls for safety goals and monitoring the goals to ensure safety [19]. Health monitors need to be implemented to observe the degradation of a circuit due to wearout. They may detect a failure or an imminent failure. The former is called diagnosis and the latter is called prognosis. The output of the monitor is subsequently analyzed to plan actions to avoid a hazard.

Wearout monitoring techniques are developed with various applications in mind. Its three main applications are: 1) Process characterization, 2) Performance-reliability optimization, and 3) Run-time health monitoring. The requirements for the run-time health monitoring application are different than for the others. For example, external measurement equipment is used for process characterization but building a run-time monitoring system based on external equipment is not pragmatic. Many wearout monitoring techniques have been investigated. This paper systematically categorizes the prominent techniques and assesses each category for the health monitoring application.

The rest of the paper is organized as follows. Section II briefly discusses the major wearout mechanisms. The most prominent monitoring techniques are categorized and assessed in Section III. The paper is concluded in Section IV.

II. SEMICONDUCTOR WEAROUT

A. Major wearout mechanisms

The Negative Bias Temperature Instability (NBTI), Positive Bias Temperature Instability (PBTI), Hot Carrier

Injection (HCI), and Time-Dependent Dielectric Breakdown (TDDB) are cited as major wearout mechanisms [3][4][7]. NBTI affects PMOS devices when the negative gate-source bias voltage draws carriers from the channel into the gate dielectric. This breaks the Si-H bonds at the silicon-oxide interface and causes a positive shift in the absolute value of the threshold voltage. The effect is more pronounced at higher temperatures. PBTI affects the NMOS devices and causes a threshold voltage shift due to the application of a positive bias voltage. It becomes critical for devices fabricated in newer process technologies that employ high-K dielectric materials [7]. HCI takes place as transistors switch and some carriers gain a high kinetic energy. These so-called hot carriers may be injected into the gate oxide and get trapped there. Hence the device degrades by way of a degrading threshold voltage and saturation current [7]. TDDB is becoming an increasing concern as the gate dielectric thickness is scaled down to the nanometer range [4]. Since the supply voltage is not scaled as aggressively as the device dimensions, a stronger electric field is formed across the gate oxide. This degrades the dielectric material and causes a conductive path through the gate dielectric layer.

B. Causes of Wearout

The following factors are known to cause wearout.

- *Newer technology*: Process scaling and high-K dielectric materials contribute to PBTI and TDDB.
- *Temperature*: Temperature influences the wearout mechanisms, like NBTI.
- *Thermal cycling*: The thermal cycling contributes to the degradation of devices composed of multiple materials, such as a semiconductor IC [1].
- *Supply voltage*: Degradation depends on the supply voltage. In particular, the dynamic voltage scaling technique contributes to device degradation.
- *Amount of workload*: The capacitive load driven by a circuit has a correlation with the degradation [15][16].
- *Switching activity*: More switching leads to an increase of hot carriers, which accelerates the HCI.

C. Indicators of Wearout

Semiconductor wearout is a gradual process, which creates a chain of degradation events. It starts with the degradation of the threshold voltage as shown in Fig. 1. The threshold voltage degradation leads to a reduction of the drain current. As a result, signals take more time to travel through the datapath. This is indicated as the increasing circuit delay in Fig. 1.

This work was partially funded by the European CATRENE RESIST subsidy project.

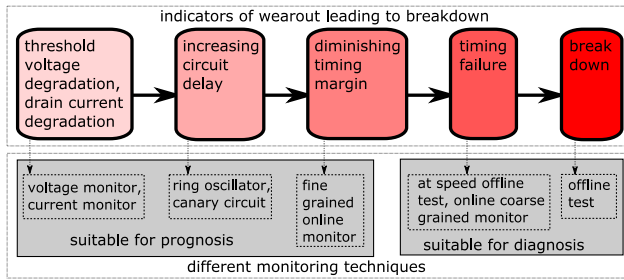


Fig. 1. Wearout monitoring using different indicators

A slight increase in the propagation delay may not immediately create a timing failure. It merely reduces the available timing margin of the datapath. At this point the circuit is still capable of functioning correctly albeit with a diminished timing margin. With the increase in stress conditions or continued aging, the timing margin diminishes to a point where a timing failure starts to occur. The degraded circuit may still function correctly at a slower clock speed and/or with relaxed operating conditions. However, a timing failure will occur at the normal operating conditions. As the aging process continues, the device eventually breaks down. At this point the circuit is not capable of functioning any longer. Different monitoring techniques to detect these indicators are discussed in the following section.

III. WEAROUT MONITORING TECHNIQUES

We first categorize the monitoring techniques based on where the function is implemented. We then categorize them further based on how the function is implemented. This categorization is shown in Fig. 2. The techniques highlighted in gray are discussed separately in the following sub-sections.

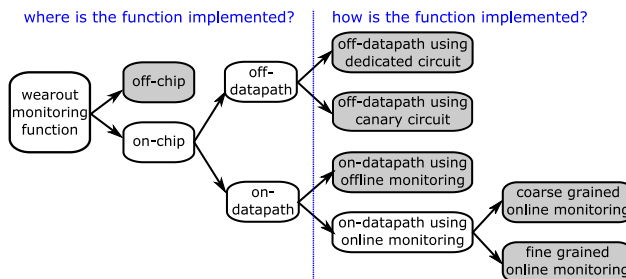


Fig. 2. categorization of different monitoring techniques

A. Off-chip monitoring

This technique applies different types of stress factors to the target die for a period of time, removes the stress, and performs an I-V measurement using external equipment. The stress factors include temperature, voltage, thermal cycling, and mechanical stress [1]. The conceptual diagram of the off-chip monitoring is shown in Fig. 3. It is typically performed for process characterization and design optimization applications.

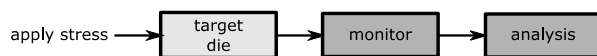


Fig. 3. Off-chip monitoring

This technique allows the extraction and analysis of a large set of measurement data on powerful computing platforms. However, it relies on external measurement equipment, which is not suitable for run-time health monitoring application. To comply with the highest safety levels defined in the ISO26262, monitors need to be implemented independent from the circuit being monitored. This is to avoid a failure common to both the circuit being monitored and the monitor. Such common cause failures may result in an undetected error. Off chip monitors are well suited to avoid a common cause failure and can be an attractive way to abide by the ISO26262 recommendations. However, for run-time health monitoring purposes the off-chip solution needs to avoid relying on external equipment.

B. Off-datapath monitoring using dedicated circuitry

The on-chip monitoring technique can be realized in two different ways. The monitoring function can observe a circuit under stress that is not part of the datapath; or it can monitor the datapath itself. The former is called off-datapath monitoring and the latter is called on-datapath monitoring. The conceptual diagram of off-datapath monitoring is shown in Fig. 4. The analysis of the monitor observations may or may not take place on-chip, hence the dotted line around the analysis block.

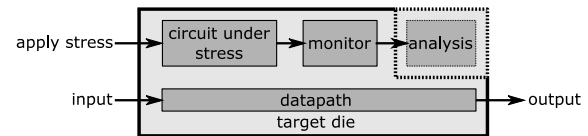


Fig. 4. Off-datapath monitoring

The circuit under stress used in off-datapath monitoring can be constructed by means of a dedicated circuit, or a modified version of the datapath. This section discusses the off-datapath monitoring by a dedicated circuit, e.g., a ring oscillator. A stress voltage is applied to stress the ring oscillator and the difference of its oscillator period before and after the stress is used to measure the performance degradation due to NBTI [2]. This technique, however, cannot eliminate the common mode environmental variation effects and cannot determine which portion of the frequency degradation is due to NBTI, and which portion is due to environmental variations. To overcome this limitation, a differential frequency measurement technique is presented in [3]. They employ a pair of ring oscillators, out of which one is stressed and the other is not. Karl et al. rely on a PMOS device that controls the current supplied to a ring oscillator [5]. The NBTI effect degrades the threshold voltage of the PMOS device, causing the current to the ring oscillator to decrease. As a result, its oscillation frequency reduces.

TDDB causes the formation of a conductive path in the gate dielectric layer. The circuit in [4] monitors a progressive decrease in the gate resistance, which is indicative of TDDB.

There are monitors that can detect multiple wearout mechanisms. The authors in [6] have presented an on-chip aging sensor to monitor NBTI and HCI related degradation. Their technique is based on measuring the threshold voltage difference between an NBTI/HCI stressed device and an NBTI/HCI unstressed device using an inverter chain. Keane et al. have presented an all-in-one on-chip solution to monitor the

aging effects due to BTI, HCI, and TDDB [7]. Their solution extends on work presented in [3], and the technique is capable of separately monitoring the effects due to these three crucial degradation mechanisms.

This technique is suitable for process characterization and design optimization. However, there are two major problems related to the health monitoring application:

1. It is difficult to match the stress applied to the dedicated circuit to the stress experienced by the datapath.
2. Even if realistic stress is applied to the dedicated circuit, it is difficult to extrapolate the results to assess the actual degradation of the functional datapath.

C. Off-datapath monitoring using canary circuit

Instead of a dedicated circuit, the circuit under stress can be realized using a replica datapath. The replicated path, known as the canary circuit, is designed such that it degrades faster than the actual datapath. For example, the V_{th} degradation can be induced by changing the body bias, oxide thickness variation, and different W/L ratio. We can replicate a few timing-critical paths and monitor their aging [8]. Based on the observed degradation of the replicated critical paths, it is possible to assess the health of the actual datapath. Three major limitations of this technique are:

1. Aging depends on the field usage. Therefore, a path that was non-critical after fabrication may become critical over the lifetime due to its operating point, workload, and its process corner [15][16][17]. Hence we may need to construct a large number of replicas to include the paths that are likely to become critical over the lifetime.
2. The replica and the actual datapath may degrade differently due to on-die process variations and on-chip temperature variations.
3. It is difficult to match the workload of the replica path to the workload of the actual datapath.

D. On-datapath offline monitoring

On-datapath monitoring addresses the limitations of off-datapath monitoring by observing the datapath itself. Based on when the datapath is observed, this technique is divided into off-line monitoring and on-line monitoring. The conceptual diagram of on-datapath offline monitoring is shown in Fig. 5.

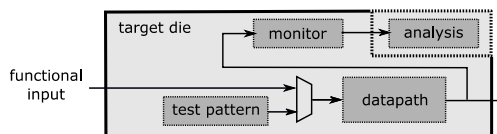


Fig. 5. On-datapath offline monitoring

Here, the datapath is monitored when it is not in the functional mode. The datapath is excited by test patterns and the test response is collected to analyze if any failure has occurred. Structural tests for automotive In-Vehicle Network (IVN) components have been presented in [9][10]. They are able to diagnose stuck-at faults or wired-and faults. Based on the test speed, this technique can be divided into two classes.

1. *Normal offline monitoring*: Test patterns are applied at the test speed, which is typically lower than the functional speed. Therefore, it can detect a breakdown situation. However, it cannot detect the timing failure that occurs when the circuit runs at the functional speed.
2. *At-speed offline monitoring*: Test patterns are applied at the functional speed. This technique can detect the timing failures caused by an increasing circuit delay.

Here we observe the degradation of the datapath itself, and thereby avoid the inaccuracy caused by the extrapolation needed for the off-datapath monitoring technique. Since this technique provides a pass/fail result, it is limited to diagnosis purposes. It cannot make a prognosis of an imminent failure.

E. On-datapath coarse grained online monitoring

Online monitoring observes the datapath when it is in the functional mode and exercised by functional input signals. The conceptual diagram is shown in Fig. 6. The online monitoring technique observes the slack of the datapath considering its actual workload and on-chip effect. As the datapath degrades with aging, the slack of the datapath between registers gradually diminishes. Based on the resolution of the online slack monitoring, this technique can be sub-divided into two categories, namely, coarse-grained online monitoring and fine-grained online monitoring.

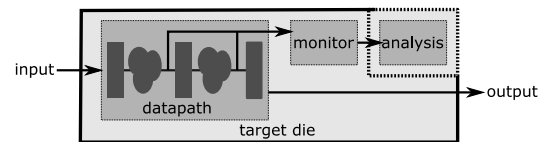


Fig. 6. On-datapath on-line monitoring

For coarse grained monitoring, the slack is monitored to yield a pass/fail indication. Different datapath slack monitoring techniques have been reported in literature [11][12][13]. A Razor flip-flop augments a delay-critical flip-flop with a shadow latch controlled by a delayed clock. By comparing the values stored in the flip-flop and the shadow latch, an error signal is generated. In case of an error, the value in the latch is utilized. By letting the error happen and handling it, they eliminate the need for the voltage margin that is otherwise necessary for “always-correct” operation in traditional designs. This technique is used to tune the voltage and clock frequency. To avoid the delayed clock, Sato et al. use a canary flip-flop that has an added delay on its data input [13]. Their canary flip-flop experiences a timing error before the main flip-flop does.

This technique does not require a dedicated circuit to stress, any stress signal, nor an extrapolation of observation. Since it only provides a pass/fail indication, the monitor can tell us if there is enough slack to meet the timing or not. In this way it provides an indication of a timing failure, but does not provide an indication of an imminent failure. As a result, it is suitable for diagnosis purposes, but not for prognosis purposes.

F. On-datapath fine grained online monitoring

Fine grained datapath monitoring measures the slack with a higher resolution to detect the gradual decrease of the available

slack and the diminishing timing margin. This technique can be realized by the triple-latch monitoring technique, which is developed to tune the clock speed [14]. This technique samples the datapath three times with a small delay interval and compares the three samples. This comparison is used to make sure that the clock speed is tuned optimally and the correct value is propagated to the next stage.

The resolution of the slack monitoring can be improved by using more than three samples. Drake et al. have presented two versions of online slack monitoring [15][16]. In [15] they describe a delay chain consisting of inverters. The delay chain is connected to the endpoint of a datapath. When an edge enters the chain, its position in the chain is captured with the rising edge of a clock. They measure the slack by looking at how far the signal edge has propagated into the delay chain. In [16], they implement the delay chain using buffers, instead of inverters. They use their technique to adjust the voltage or frequency of a microprocessor to save energy during low-temperature and low-activity periods. Blome et al. describe a chain of delay buffers to monitor propagation latency [18]. With the arrival of the clock, some registers will capture the correct value, while the others will store an incorrect value. An XOR operation of each of the delayed registers with the correct value produces a bit vector that shows the propagation delay.

Agarwal et al. describe a circuit failure prediction concept, which enables close to best-case design instead of traditional worst-case design [17]. Their monitoring circuit is based on the concept of stability checking during the guard-band interval by detecting signal transitions too close to the capturing clock. This is referred to as the guard-band violation. A guard-band violation means that one or more paths have aged enough to creep into the guard-band interval.

The fine grained online monitoring technique observes the diminishing timing margin of aging critical paths, considering the effect of process variation, ambient conditions, and workload. However, monitoring the large number of datapaths of a typical design is not pragmatic. Therefore, monitors are designed to observe a few aging critical datapaths. However, aging depends on the field usage. Therefore, finding the aging critical datapaths during the design-time is challenging.

IV. CONCLUSION

We categorize the available monitoring techniques and qualitatively assess their value for health monitoring. We have found that off-chip monitoring is “suitable by design” to eliminate common cause failures and can help fulfill the ISO26262 recommendations. However, their dependency on external measurement equipment needs to be avoided.

We find that the on-datapath techniques are particularly suitable for health monitoring. They need to be implemented independently from the circuit being monitored, for example by means of independent power supply, oscillator, etc. The following two techniques deserve further attention:

1. *On-datapath offline monitoring* becomes interesting when at speed tests are conducted. Since it operates at the functional speed, we can detect timing failures. Logic-BIST circuits can be used to realize at-speed monitoring.

2. *On-datapath fine grained online monitoring* is an interesting choice as it measures timing margins with higher resolution. However, finding the aging critical paths during the design-time remains a problem. Design-time analysis of critical path delays are performed using library cells whose delays are characterized for an output load and input slew with respect to a given PVT corner. To find out the path that would become critical during the lifetime of the chip we need an aging-aware library whose cells are characterized for load, slew and aging.

REFERENCES

- [1] M. Baybutt et al., “Improving Digital System Diagnostics Through Prognostic and Health Management (PHM) Technology,” *IEEE Trans. Instrumentation and Measurement*, vol. 58, no. 2, pp.255-262, Feb. 2009
- [2] V. Reddy et al., “Impact of negative bias temperature instability on digital circuit reliability,” in *Proc. IEEE Int. Reliability Physics Symp.* pp. 248-254, 2002.
- [3] T.H. Kim et al., “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.
- [4] J. Keane et al., “An Array-Based Test Circuit for Fully Automated Gate Dielectric Breakdown Characterization,” *IEEE Trans. Very Large Scale Integration Systems*, vol. 19, no. 5, pp. 787-795, May 2011.
- [5] E. Karl et al., “Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation,” in *IEEE Int. Solid-State Circuits Conf.*, pp. 410-411, 2008.
- [6] K.K. Kim, W. Wang, and K. Choi, “On-Chip Aging Sensor Circuits for Reliable Nanometer MOSFET Digital Circuits,” *IEEE Trans. Circuits and Systems*, vol., 57, no., 10, pp. 798-802, Oct. 2010.
- [7] 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.
- [8] J.W. Tschanz et al., “Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage,” *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396-1402, Nov. 2002.
- [9] A. Cook et al., “Structural In-Field Diagnosis for Random Logic Circuits,” in *Proc. IEEE European Test Symp.*, pp. 111-116, 2011.
- [10] U. Abelein et al., “Non-intrusive integration of advanced diagnosis features in automotive E/E-architectures,” in *Proc. Design, Automation and Test in Europe Conf.*, pp. 1-6, 2014.
- [11] D. Ernst et al., “Razor: A low-power pipeline based on circuit-level timing speculation,” in *Proc. IEEE Int. Symp. Microarchitecture*, pp. 7-18, 2003.
- [12] S. Das et al., “A self-tuning DVS processor using delay-error detection and correction,” *IEEE J. Solid-State Circ.*, vol. 41, no. 4, pp. 792-804, Apr. 2006.
- [13] T. Sato and Y. Kunitake, “A Simple Flip-Flop Circuit for Typical-Case Designs for DFM,” in *Proc. IEEE Int. Symp. Quality Electronic Design*, pp. 539-544, 2007.
- [14] T. Kehl, “Hardware self-tuning and circuit performance monitoring,” in *Proc. IEEE Int. Conf. Computer Design*, pp. 188-192, 1993.
- [15] A. Drake et al., “A Distributed Critical-Path Timing Monitor for a 65nm High-Performance Microprocessor,” in *Proc. IEEE Solid-State Circ. Conf.*, pp. 398-399, 2007.
- [16] C.R. Lefurgy et al., “Active Guardband Management in Power7+ to Save Energy and Maintain Reliability,” *IEEE Micro*, vol. 33, no. 4, pp. 35-45, 2013.
- [17] M. Agarwal et al., “Circuit Failure Prediction and Its Application to Transistor Aging,” in *IEEE VLSI Test Symp.*, pp. 277-286, 2007.
- [18] J. Blome et al., “Self-calibrating Online Wearout Detection,” in *Proc. IEEE Int. Symp. on Microarchitecture*, pp. 109-122, 2007.
- [19] Road Vehicles Functional Safety Standard, ISO 26262, [Online], <http://www.iso.org>