Dependable Design in Nanoscale CMOS Technologies – Challenges and Solutions

Vikas Chandra
ARM R&D
vikas.chandra@arm.com

***Abstract*—With nearly three decades of continued CMOS scaling, the devices have now been pushed to their physical and reliability limits. Imperfections in manufacturing are unavoidable due to atomistic scale of nanoscale devices. With technology scaling, early life failures are becoming increasingly common due to random manufacturing defects and SEUs are of great concern during the normal lifetime of the design. Designs manufactured correctly will wearout and become unreliable over time because of mechanisms like NBTI and gate oxide breakdown. The impact of unreliability results in time-dependent variability, where the electrical characteristics of the devices will vary statistically in a temporal manner, directly translating into timing uncertainty in manufactured chips. Further, dynamic variations in voltage and temperature will induce variability in the performance of the design. Design techniques which can adapt to these variations can make the design more resilient by mitigating errors on the fly. The three components of adaptive design techniques are failure prediction, failure detection and failure recovery. Adaptive design techniques with particular emphasis on error-tolerant techniques will be reviewed.**

I. INTRODUCTION

Designs in scaled CMOS technologies are inherently unreliable due to spatial, temporal and dynamic variations [2], [7]. The spatial variations are due to manufacturing imperfections which result from mechanisms like variations in lithography and natural randomness in channel dopant placement. The temporal variations in the design are due to early life failures, transient errors and the wearout of the devices. The dynamic variations are due to fluctuations in supply voltage and temperature which can impact the timing behavior of the design. Adaptive design techniques are required to make the design more resilient against these variations. The rest of the paper discusses these variability mechanisms in detail. Various adaptive design techniques needed to make the design more dependable and resilient are also reviewed.

II. SPATIAL VARIATIONS

There are almost innumerable ways in which nature introduces imperfections into chip manufacturing. We will discuss some variation sources that are starting to have a direct impact on the electrical yield of circuits. These can be classified into two groups, based on the mechanism of the variation:

- Systematic, process and apparatus induced variations: These are induced directly because of apparatus imperfection and systematic design characteristics. For example, variations in optical focus and exposure dose, mask overlay, etch rates and chemical mechanical polishing.
- Random variations: These are induced because of inherent randomness in the process. For example, random dopant fluctuation (RDF) and stress-induced mobility variation.

III. TEMPORAL VARIATIONS

Due to shrinking geometry, the reliability of nanoscale devices decreases substantially. Figure 1 shows the failure rate of designs over time. The curve (known as bathtub curve due to its shape) has three distinct regions.

- Infant mortality : Also known as early life failures. The high failure rate is due to random manufacturing defects like poor contacts/vias, gate-to-source shorts etc.
- Normal lifetime : The failures are due to transient errors (for example, single event upsets) and failure rate is fairly constant in this region
- Wearout/Aging: The sharp increase in the failure rate is due to aging mechanisms like NBTI and gate oxide breakdown [6], [8].

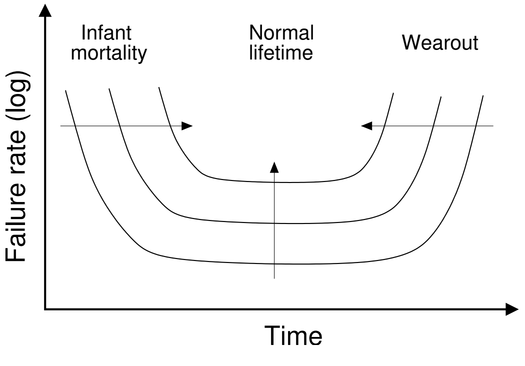

Fig. 1. Failure rate during the design life-cycle

Due to the increase in unreliability of devices in nanoscale technologies, the bathtub curve is shrinking as shown in Figure 1. The rate of infant mortality is going up, the designs are becoming more susceptible to transient errors and the onset of wearout may even happen during the device lifetime.

IV. DYNAMIC VARIATIONS

The two main sources of dynamic variations are:

- Voltage fluctuations: Abrupt changes in die-level switching activity can cause large current transients in the power delivery system, resulting in Vdd droop and overshoot fluctuations. The magnitude and duration of Vdd droops and overshoots depend on the interaction of capacitive and inductive parasitics at the board, package, and die levels with changes in current demand.
- Temperature variations: The variations in temperature depend on workload, environmental conditions, and the heat-removal capability of the package.

Conventional microprocessor performance reduces as Vdd decreases or as temperature increases [3]. These dynamic variations can cause timing errors in the design and impact functionality.

V. DESIGN MARGINS

One way to design around these variations is to build static margins in the design. However, margins (or guard-bands) leave too much performance on the table and are becoming prohibitively expensive with technology scaling. Since the variation mechanisms described above vary statistically, it is not efficient to build static margins in the design because of power, performance and area overhead. A more efficient approach is build designs which are adaptive and hence error tolerant.

VI. ADAPTIVE DESIGN TECHNIQUES

Adaptive design techniques will become necessary to design resilient system which can cope with spatial, temporal and dynamic variations. The errors should be detected and corrected in a transparent manner to the user. In addition, the adaptive design should work at the most efficient point for all conditions and should respond by sensing the variations. The typical adaptive knobs in a design are frequency, supply voltage and body biasing. To build a resilient system, support is required at the circuit, micro-architecture and system level. The three aspects of an adaptive system are:

- Failure prediction: The main idea it to detect slow-changing variations before they cause errors. The system can respond by changing voltage, frequency etc.
- Failure detection: When an error occurs, it is important to detect them quickly. The detection method could be implemented with software (assertion, bounds checking), redundancy (spatial, temporal), coding (parity, residue) etc.
- Failure recovery: When an error is detected, there are multiple options to recover - local recovery (fast, high circuit complexity), instruction replay or checkpointing with roll-back (slow, high software complexity). The choice to use any of these methods depends on the requirements on power, performance and area.

A. *Error tolerant architectures*

As process technology scales, the local variations worsen thereby undermining the efficacy of traditional adaptive techniques. Error tolerant techniques address local variations by scaling voltage and frequency until the point where the processor incurs timing errors. Error-detection circuits flag such an occurrence and engage a recovery mechanism to restore correct state. This eliminates all worst-case safety margins and enables significantly improved performance and energy efficiency over the traditional techniques. Their relative complexity makes the general applicability of such systems difficult. However, they are naturally amenable for communications and signal-processing where existing mechanisms can be overloaded to detect and correct timing errors [1], [4], [5], [9].

VII. CONCLUSIONS

Reliability challenges in CMOS technologies are increasing due to dominance of spatial, temporal and dynamic variations in the nanoscale era. The spatial variations are due to the atomistic scale of the manufacturing process, the temporal variations are due to the drift in the device parameters over time and the dynamic variations are due the fluctuations in supply voltage and temperature. The dependability of design in the presence of these variations becomes very challenging. Adaptive design techniques are needed to react to the failures and correct errors on the fly. In particular, error tolerant architectures will be needed to make the design resilient against these variations.

ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Shidhartha Das, Dr. Krisztian Flautner and Dr. Rob Aitken for various insightful discussions on the topic.

REFERENCES

[1] D. Blaauw *et al*, “RazorII: In-Situ Error Detection and Correction for PVT and SER tolerance,” *IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2008.

[2] S. Borkar *et al*, “Design and reliability challenges in nanometer technologies,” *IEEE Design Automation Conference*, 2004.

[3] K. A. Bowman *et al*, “Energy-Efficient and Metastability-Immune Resilient Circuits for Dynamic Variation Tolerance,” *IEEE Journal of Solid-State Circuits*, Vol. 44, No. 1, pp. 49-63, Jan 2009.

[4] S. Das *et al*, “A Self-Tuning DVS Processor using Delay-Error Detection and Correction,” *IEEE Journal of Solid-State Circuits*, pp. 792-804, Apr. 2006.

[5] R. Hegde *et al*, “A voltage overscaled low-power digital filter IC,” *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 2, pp. 388-391, Feb 2004.

[6] B. Kaczer *et al*, “Impact of MOSFET gate oxide breakdown on digital circuit operation and reliability,” *IEEE Transactions on Electron Devices*, Vol. 49, No. 3, pp. 500-506, Mar 2002

[7] J. W. McPherson, “Reliability Challenges for 45nm and Beyond,” *IEEE Design Automation Conference*, 2006.

[8] B. C. Paul *et al*, “Impact of NBTI on the temporal performance degradation of digital circuits,” *IEEE Electron Device Letters*, Vol. 26, No. 8, pp. 560-562, Aug 2005.

[9] F. Worm *et al* “A robust self-calibrating transmission scheme for on-chip networks,” *IEEE Transactions on VLSI Systems,* Vol. 13, No. 1, pp. 126-139, Jan. 2005.