What happens when circuits grow old: Aging issues in CMOS design

Sachin S. Sapatnekar
ECE Department, University of Minnesota
200 Union St. SE
Minneapolis, MN 55455, USA

ABSTRACT

As CMOS technologies have shrunk to tens of nanometers, aging problems have emerged as a major challenge. There has been tremendous progress in developing new methods for modeling and diagnosing reliability at the level of individual transistors, but much less work on propagating these models to higher levels of abstraction to analyze and optimize the reliability of larger circuits. This talk will provide an introduction to various circuit aging mechanisms and will then discuss research that develops computer-aided design techniques for estimating and enhancing the reliability of large digital circuits, examining solutions that could practically be applied to analyze or improve the lifetime of a design while maintaining consistency to accurate device-level models and the associated physics.

CIRCUIT-LEVEL ANALYSIS OF AGING

Aging effects are described by the classical bathtub curve: after a steep initial failure rate, the failures level off for a while before rising again, resulting in a bathtub-like shape for the number of failures as a function of time. Aging is strongly sensitivity to process parameter variations, as well as to on-chip temperature and supply voltage level changes. Here, we overview analysis methods for prominent aging mechanisms, based on both conventional and newer physical models.

Bias temperature instability (BTI) is a device aging phenomenon that causes threshold voltage shifts over long periods of time in the presence of voltage stress at the gate, eventually causing the circuit to fail to meet its specifications. A PMOS transistor in an inverter experiences negative BTI stress when its gate node is at logic 0, and the resulting increase in the threshold voltage is partially reversed when the voltage stress is removed (i.e., a logic 1 is applied). A similar phenomenon of positive BTI affects the threshold voltage of NMOS devices when they are stressed, and relaxes the degradation on the removal of stress. The magnitude of degradation depends on the ratio of the stressed time to unstressed time, i.e., the signal probability (SP) of the gate input. There are two widely-used theories for BTI. The reaction-diffusion (R-D) model [1,2] explains the degradation due to the accumulation of positive charges as Si-H bonds at the interface are broken and hydrogen diffuses away; the removal of this stress allows partial restoration of these bonds. The charge-trapping (CT) model [3] provides an alternative explanation, where defects in gate dielectrics can capture charged carriers, causing the threshold voltage to degrade. Neither theory fully explains experimental observations, and it has been posited that R-D and CT mechanisms coexist. The CT model predicts high variability in small devices, making variability a major factor for memories. For logic circuits, however, large transistors and averaging effects on critical paths significantly mitigate variability [4].

Hot carrier injection (HCI) in MOSFETs is caused by the acceleration of carriers (electrons/holes) under lateral electric fields in the channel, to the point where they gain enough energy and momentum to cause damage, degrading mobilities and threshold voltages. Hot carrier effects cumulatively build up over prolonged periods, causing the circuit to age, resulting in irreversible degradations that may lead to circuit failure.

At the device level, HCI is modeled by energy-driven theories that generalize the ideas of the traditional lucky electron model [5], and explain the mechanism of carrier-induced degradation for short-channel devices at low supply voltages. The energy-driven framework [6] includes the effects of electrons of various energies, from high-energy channel hot carriers (CHCs) to low-energy channel cold carriers (CCCs).

Existing circuit-level work on HCI is largely based on the lucky electron model. The approach in [7] applies a duty factor to capture the effective stress time for HC effects, which assumes constant HC stress during signal transitions and models the duty factor to be proportional to the transition time, and also simultaneously incorporates BTI effects. Energy driven models are leveraged in [8] uses the new energy-driven theories described above, and defines an age gain per transition using quasistatic characterization. Using abstractions based on the signal probability (SP) and transition density (TD), this is used to build a methodology for timing analysis of large-scale circuits.

The HCI rate increases as $t^{1/2}$, where t is the time variable. Since the multiplicative constant is relatively small, in the short term, HCI is overshadowed by BTI effects, which vary as t^n, for $n \approx 0.1$–0.2, but with a larger constant multiplier.

Time-dependent dielectric breakdown (TDDB) in gate oxides is an irreversible reliability phenomenon that results in a sudden discontinuous increase in the conductance of the gate oxide at the point of breakdown, as a result of which the current through the gate insulator increases significantly [9]. The time to breakdown can be modeled statistically using a Weibull distribution. Circuit-level analysis approaches for TDDB are limited. Memory effects have been addressed in [10]. For logic circuits, a conventional area-scaling based method is presented in [11]. However, it has been shown that logic circuits are inherently resilient to variation [12,13], and the area-scaling model is excessively pessimistic.

Electromigration failures may result when a current flows through an on-chip wire over a long period of time, causing a physical migration of atoms in the wire, particularly near vias in copper interconnects [14] and in a statistical manner [15]. The electron wind leads to a net flux of metal atoms, creating

voids upstream and hillocks downstream at locations of atomic flux divergence. The migration causes a reverse stress known as the Blech effect that resists electromigration. Electromigration is witnessed most notably in power and ground wires, where current flow is mostly unidirectional, but AC electromigration is also seen in signal wires.

CIRCUIT OPTIMIZATION

The analysis models above can be used at various levels of abstraction for circuit optimization. In this area, research at the transistor and gate levels has focused primarily on BTI and to some extent, HCI, although in many cases, the models that are used do not reflect contemporary reality. Gate oxide TDDB optimization work is very limited: the research in [12] shows how selective upsizing can make a circuit more robust to such errors. For interconnect reliability, EM optimization is widely used in mainstream interconnect optimization. The primary strategy is based on Black's law models and limits the current density on a wire to a specified threshold. However, newer models are sparsely represented in the optimization literature: e.g., [15] shows the inherent resilience of power grids to electromigration in copper.

BTI optimization techniques can largely be classified into a few categories. One class of methods attempts to rebalance the signal probabilities in a circuit to even out the wearout on pull-up and pull-down paths in a circuit: such methods have been applied to memory and logic. An alternative approach temporarily deactivates memory units on a rotating basis to allow for recovery from wearout [16].

A cross-layer system optimization method is presented in the GNOMO approach [17]. The approach is based on observing that at higher supply voltage, a computation completes faster: this saved time is used to put the circuit to sleep (which allows recovery). At isoperformance, this results in lower aging, which translates to power savings through reduced delay margining, as compared to conventional design.

Static presilicon [18] or adaptive postsilicon [19,20] techniques may also be used to pad the circuit against BTI degradation based on time sensors, history sensors that track usage patterns, or surrogate sensor circuits. Optimization knobs include changing the supply voltages, body biases, or both, using using dynamic cooling, and through redundancy and disposable cores. Adaptive sense/respond mechanisms require sensing capabilities: related work includes "silicon odometers" and similar sensors to capture BTI, but also HCI and TDDB [21]. Other proposed ideas include online test techniques for failure analysis, introspective sensor-based methods for reduced aging, and intelligent power gating.

Finally, circuits may be sent to sleep states that are designed to minimize BTI degradation through input vector control [22]: however, the gains of such methods are relatively small.

REFERENCES

[1] M. A. Alam and S. Mahapatra, "A comprehensive model of PMOS NBTI degradation," *Microelectronics Reliability*, 45(1), pp. 71–81, Jan. 2005.

[2] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "A finite-oxide thickness-based analytical model for negative bias temperature instability," *IEEE Transactions on Device and Materials Reliability*, 9(4), pp. 537–556, December 2009.

[3] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P.-J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, "Recent advances in understanding the bias temperature instability", *Proceedings of the IEEE International Electronic Devices Meeting*, pp. 4.4.1–4.4.4, 2010.

[4] J. Fang and S. S. Sapatnekar, "Understanding the impact of transistor-level BTI variability," *Proceedings of the IEEE International Reliability Physics Symposium*, pp. CR2.1–CR2.6, 2012.

[5] S. Tam, P.-K. Ko, and C. Hu. "Lucky-electron model of channel electron injection in MOSFET's," *IEEE Transactions on Electron Devices*, D-31(9), pp. 1116–1125, September 1984.

[6] C. Guerin, V. Huard, and A. Bravaix. "The energy-driven hot-carrier degradation modes of nMOSFETs," *IEEE Transactions on Device and Material Reliability*, 7(2), pp. 225–235, 2007.

[7] D. Lorenz, G. Georgakos, and U. Schlichtmann, "Aging analysis of circuit timing considering NBTI and HCI," *Proceedings of the IEEE International On-Line Testing Symposium*, pp. 3–8, 2009.

[8] J. Fang and S. S. Sapatnekar, The impact of hot carriers on timing in large circuits," *Proceedings of the Asia-South Pacific Design Automation Conference*, pp. 591–596, 2012.

[9] E. Y. Wu, E. J. Nowak, A. Vayshenker, W. L. Lai, and D. L. Harmon, "CMOS scaling beyond the 100-nm node with silicon-dioxide-based gate dielectrics," *IBM Journal of Research & Development*, 46(2/3), pp. 287–298, March–May 2002.

[10] F. Ahmed and L. Milor, "Analysis and on-chip monitoring of gate oxide breakdown in SRAM cells," *IEEE Transactions on VLSI Systems,* 20(5), pp. 855–864, May 2012.

[11] K. Chopra, C. Zhuo, D. Blaauw, and D. Sylvester, "A statistical approach for full-chip gate-oxide reliability analysis," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 698–705, 2010.

[12] J. Fang and S. S. Sapatnekar, "Scalable methods for the analysis and optimization of gate oxide breakdown," *Proceedings of the International Symposium on Quality Electronic Design*, pp. 638–645, 2010.

[13] J. Fang and S. S. Sapatnekar, "Accounting for inherent circuit resilience and process variations in analyzing gate oxide reliability," *Proceedings of the Asia-South Pacific Design Automation Conference*, pp. 689–694, 2011.

[14] S. M. Alam, C. L. Gan, F. L. Wei, C. V. Thompson, and D. E. Troxel, "Circuit-level reliability requirements for Cu metallization," *IEEE Transactions on Device and Materials Reliability*, 5(3), pp. 522–531, 2005.

[15] V. Mishra and S. S. Sapatnekar, "The Impact of Electromigration in Copper Interconnects on Power Grid Integrity," *Proceedings of the ACM/EDAC/IEEE Design Automation Conference*, 2013.

[16] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Impact of NBTI on SRAM read stability and design for reliability," *Proceedings of the International Symposium on Quality Electronic Design* pp. 210–218, 2006.

[17] S. Gupta and S. S. Sapatnekar, "GNOMO: Greater-than-NOMinal Vdd Operation for BTI mitigation," *Proceedings of the Asia-South Pacific Design Automation Conference*, pp. 271–276, 2012.

[18] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "NBTI-Aware Synthesis of Digital Circuits," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 370–375, 2007.

[19] S. V. Kumar, C. H. Kim, and S. S. Sapatnekar, "Adaptive techniques for overcoming performance degradation due to aging in digital circuits," *Proceedings of the Asia-South Pacific Design Automation Conference*, pp. 284–289, 2009.

[20] E. Mintarno, J. Skaf, R. Zheng, J. Velamala, Y. Cao, S. Boyd, R. W. Dutton, and S. Mitra, "Optimized self-tuning for circuit aging," *Proceedings of Design, Automation and Test in Europe*, pp. 586–591, 2010.

[21] J. Keane, X. Wang, D. Persaud, and C. H. Kim, "An all-in-one silicon odometer for separately monitoring HCI, BTI, and TDDB," *IEEE Journal of Solid-State Circuits*, 45(4), pp. 817–829, April 2010.

[22] Y. Wang, X. Chen, W. Wang, V. Balakrishnan, Y. Cao, Y. Xie, and H. Yang, "On the efficacy of input vector control to mitigate NBTI effects and leakage power," *Proceedings of the International Symposium on Quality Electronic Design*, pp. 19–26, 2009.