Threshold Voltage Variation Effects on Aging-Related Hard Failure Rates

Brian Greskamp Smruti R. Sarangi Josep Torrellas

University of Illinois at Urbana-Champaign
http://iacoma.cs.uiuc.edu

Abstract— **This paper quantifies the impact of threshold voltage variation on aging-related hard failure rates in a high-performance 65nm processor. Simulations show that threshold voltage variations can accelerate aging substantially, depending on the thermal resistance of the heatsink and the total leakage power of the processor before variation. For unfavorable values of these parameters, our models suggest that the time at which 1% of the processors have failed can decrease by about 60%.**

I. Introduction

Process variation is known to have a significant effect on transistor power consumption. Variations in threshold voltage (V_t) are of special interest because they are projected to have one of the highest variation ranges [1]. Moreover, they strongly affect a transistor's subthreshold leakage power — since leakage power is exponentially dependent on the threshold voltage. With current high-performance processors pushing the limits of a package's power dissipation capability, any excess power is a concern.

Meanwhile, industry is focusing increased attention on failures due to device aging [2]. While current industry designs have a design lifetime of seven to ten years [3], [4], guaranteeing device lifetimes becomes more difficult in smaller technologies. Any further decreases in lifetime reliability could lead to a dramatic increase in the number of failures in the field. Consequently, it is necessary to take careful account of any additional threats to reliability.

Failure rates for many important aging mechanisms are exponentially dependent on temperature [5]. At the same time, threshold voltage variation, by inducing significant increases in leakage, can elevate die temperatures. This is because leakage power often comprises up to 50% of total power consumption in a processor [6]. A reasonable question, therefore, is whether threshold voltage variation itself, through temperature, has a major impact on lifetime reliability.

This paper specifically examines this question. It shows that threshold voltage variation may affect the lifetime reliability of a processor significantly. Depending on the thermal resistance of the heatsink and the total leakage power of the processor before variation, the time at which 1% of the processors have failed can decrease by about 60%.

This paper proceeds as follows. It first presents models of lifetime reliability, power, temperature, and threshold voltage variation — all derived from previous work. It then uses them to evaluate the impact of threshold voltage variation on the reliability of a single unit in a die and of a high-performance processor.

II. Models

A. Lifetime Reliability Model

Srinivasan *et al.* [7] model four different hard-failure mechanisms and give their mean time to failure (MTTF) expressions. The failure mechanisms are: (1) Electromigration, where electrons flowing through a wire displace metal atoms over time, causing opens and shorts; (2) Stress migration, with effects similar to electromigration but caused by thermo-mechanical stress; (3) Time Dependent Dielectric Breakdown (TDDB), where conductive paths form in the gate dielectric over time; and (4) Thermal cycling, where repeated heating and cooling of the chip (e.g., due to switching it on and off) causes mechanical failure of the die or package.

Degradation rates for these failure mechanisms except thermal cycling have an exponential temperature dependence governed approximately by the Arrhenius relation: the rate of degradation r is $r(T) = k_1 e^{-k_2/T}$, where k_1 and k_2 are empirically-determined constants. The degradation rate increases rapidly with temperature, and the MTTF plummets. The TDDB degradation rate also depends exponentially on the oxide thickness [5]. Previous work [8] has already identified oxide thickness variation as a key threat to reliability. For simplicity, however, our work ignores it to focus on threshold voltage variation.

To model the spatial distributions of process variation and temperature, we partition the die into a grid of approximately 1,000 rectangular cells. Within each cell, the temperature and the systematic process parameters are assumed constant. Cell failures are assumed to be independent and, for each mechanism, are modeled by assigning a lognormal Time To Failure (TTF) distribution to each cell. For each mechanism in each cell, the MTTF is determined from the equations and constants in [7]. The lognormal TTF distribution is then constructed with the following parameters, also from [7]:

$$\mu = ln(MTTF) - \sigma^2/2, \quad \sigma = 0.5 \qquad (1)$$

The MTTFs of the four failure mechanisms in a given cell are assumed to be equal to each other at 80°C. This assumption is also made in [7], and is motivated by simplicity. To see the relative MTTFs of the different failure mechanisms as temperature scales, refer to Figure 1. The figure shows the

MTTF values normalized to the values at 80°C. The *Combined* line shows the combined effect of all four failure mechanisms.

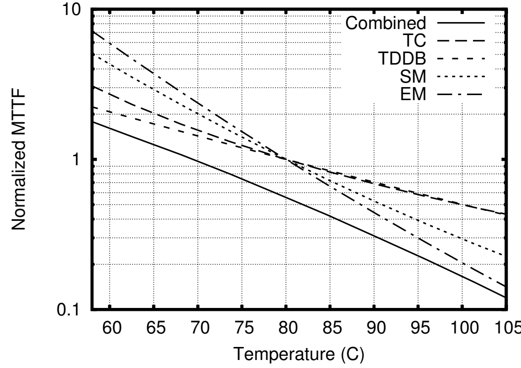

Fig. 1. Normalized MTTF for Thermal Cycling (TC), Time Dependent Dielectric Breakdown (TDDB), Stress Migration (SM), and Electromigration (EM) as a function of the temperature. The MTTF values are normalized to those at 80°C.

The MTTF is not a very practical metric for semiconductor reliability. Manufacturers are more concerned with how long they can *guarantee* a product rather than when it will fail on average. They like to make statements like "There is a 99% probability that the product will function for seven years or more" rather than "The product will last 50 years on average." Therefore, as a metric for design lifetime, this paper uses $TTF_{1\%}$, namely the time at which 1% of specimens will have failed. Similar metrics are used in industry (e.g., [4]).

B. Power and Temperature Model

Figure 2 shows an approximation of the thermal model used in this work. The die connects to the heatsink though a highly conductive heat spreader, whose purpose is to diffuse heat, ameliorating local hotspots and bringing all parts of the die to a similar temperature. Because the die is mounted face-down on the package, heat dissipated in cell i must flow through a substrate resistance R_{csi} to reach the heat spreader. Finally, the heatsink dissipates heat through thermal resistance R_{se} to the environment, which is assumed to be at 45°C. We assume steady-state operation, therefore ignoring heat storage in the thermal components. Additionally, as shown in Figure 2, we allow lateral conduction in the die by modeling a large thermal resistance between adjacent cells.

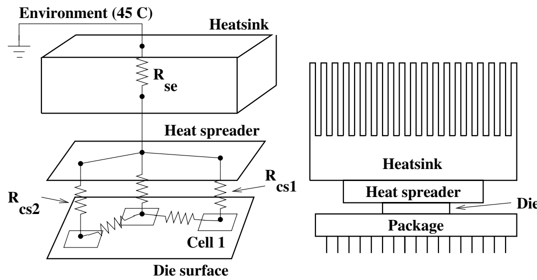

Fig. 2. Thermal assembly (right) and electrical equivalent (left) of processor die, heat spreader, and heatsink.

Total power is the sum of dynamic and static power. Dynamic power is assumed to be unaffected by process variation, while static power varies significantly. To model the latter, we use the simplified BSIM3-derived subthreshold leakage model from HotLeakage [6]. Using this model and assuming constant supply voltage, leakage current and static power (P_s) are both proportional to:

$$\begin{aligned} P_s &\propto \mu \left(\frac{kT}{q}\right)^2 exp\left(\frac{-V_t + c_2}{c_3 kT/q}\right) \\ \text{where} \quad V_t &= V_{t0} - c_1(T - T_0) \end{aligned}$$

In the formula, μ is the mobility, which is proportional to $T^{-1.5}$; V_{t0} is the threshold voltage at nominal temperature (T_0), which includes the effect of process variation; V_t is the threshold voltage at the actual operating temperature. For constants c_1, c_2, and c_3, we use the values from HotLeakage's 65nm model, namely $c_1 = 5.0 \times 10^{-4} V/K$, $c_2 = 39mV$, and $c_3 = 1.3$.

C. Variation Model

Threshold voltage variation can be die-to-die (D2D), within-die (WID) systematic, and WID random [9]. At 90nm, D2D variation was slightly greater than WID systematic variation, but the relative contribution of D2D variation is diminishing as technology scales [10]. Also, the amounts of WID systematic and WID random variation are approximately equal [1]. Consequently, this paper assumes that for 65nm, each of the three components (D2D, WID random, and WID systematic) contributes equally to the total V_{t0} variance $\sigma^2_{V_{t0}}$.

The components of V_{t0} variation are modeled as follows. We first assign a value of the WID systematic component to each cell using a multivariate normal distribution as described in [9]. This spatial distribution is generated from three intuitive statistics: the nominal threshold voltage V_{t0_nom}, the standard deviation σ_s of the WID systematic variation, and a parameter ϕ that describes the degree of spatial correlation.

Our spatial correlation model assumes position independence and anisotropy, so that the correlation function $\rho(r)$ is only a function of the distance r between two cells. We use the Spherical function [11] for $\rho(r)$, which fits Friedberg *et al.*'s [12] experimental data well. In this correlation function, cells separated by a distance of ϕ or more are totally uncorrelated. Following Friedberg *et al.*'s data, we set ϕ to be half the length of the longest side of the die.

After the WID systematic component at each cell is determined, we add to every cell a fixed offset representing the D2D variation component. This offset is normally distributed across dies with standard deviation σ_{d2d}. Finally, we add to each transistor a normally-distributed WID random component. This component has a mean of zero, a standard deviation σ_r, and no spatial correlation.

Based on projected variation levels from ITRS, we set $\sigma_{V_{t0}} = 0.09 \times V_{t0_nom}$. Using the assumption of equal contributions for the three components of variation, we set the standard deviation of each component to $(0.09/\sqrt{3}) \times V_{t0_nom} = 0.052 \times V_{t0_nom}$.

III. EXPERIMENTS

Since the MTTF is chiefly dependent on temperature, the main means by which V_{t0} variation affects reliability is by increasing the static power P_s and, therefore, the temperature of the device. Figure 3 shows the relationships between the variables. V_{t0} variation will cause some sections of some dies to have low-V_{t0} transistors. With decreasing V_{t0}, P_s increases exponentially. This increases the temperature linearly, which in turn has an exponential effect on P_s. This positive feedback loop can be seen in the figure. Under normal conditions, the heat flow out of the heatsink allows the system to reach equilibrium and thermal runaway does not occur. However, the higher temperature induces a lower MTTF.

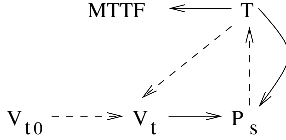

Fig. 3. Relationships between variables. Solid lines represent exponential dependences, while dashed lines are linear dependences.

In the following, we consider the effect on the MTTF at the cell level and at the processor level.

A. Per-Cell Effects

Perhaps the simplest way to understand the system in Figure 3 is by studying a single cell of an otherwise nominal chip. Because power dissipation within the cell is small compared to the total chip's dissipation, a change in the cell's leakage power has little impact on the heat spreader temperature. Therefore, we assume that the temperature of the spreader remains constant. If we call P_{s0} the static power consumed by the cell at the original operating temperature (T_0), and P_{s1} the static power consumed by the cell at a new temperature T_1, the equilibrium temperature for the cell is then given by:

$$T_1 \quad = \quad T_0 + R_{cs}(P_{s1} - P_{s0}) \qquad (2)$$

Since P_{s1} is a function of V_{t0} and T_1, the only unknown is T_1. Lacking an analytic solution, we solve Equation 2 numerically.

As an example, assume a 50mm^2 processor chip divided into 1,000 cells, and take a cell in a high-temperature unit such as the register file. We assume $T_0 = 85°C$, $R_{cs} = 100K/W$, and that the heat spreader temperature remains fixed at 70°C. If we set the threshold voltage of all the transistors in the cell to the same value V_{t0}, different from V_{t0_nom} (to model process variation to a first approximation), the T_1 and MTTF of the cell will change. Specifically, if V_{t0} is lower than V_{t0_nom}, MTTF will decrease.

Figure 4 shows the resulting cell MTTF for different values of V_{t0}/V_{t0_nom}. MTTF is normalized to its value at V_{t0_nom}. The chart shows three lines, each assuming a different value of cell leakage power P_{s0} at V_{t0_nom}.

Figure 4 can be used to see the effect of the D2D and WID systematic components of V_{t0} variation. Recall that these components affect all transistors in the cell equally,

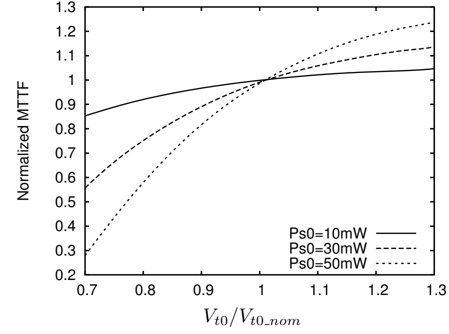

Fig. 4. Normalized MTTF of a register file cell as we vary the V_{t0} for the whole cell. The three lines correspond to different values of leakage power P_{s0} at V_{t0_nom}.

causing a cell-wide increase or decrease in V_{t0}. Given the standard deviations of the D2D and WID systematic components assumed in Section II-C, the probability of having a cell-wide value of $V_{t0} \leq 0.83 \times V_{t0_nom}$ is approximately 1%. In these conditions, the figure shows that a cell with an initial $P_{s0} = 50mW$ decreases its MTTF by 35% or more. Consequently, if unfavorable values of the D2D and WID systematic components of variation strike a cell, the cell may significantly decrease its lifetime.

In contrast, WID random variation has a lower impact. The reason is that it heats some transistors in the cell and it cools others, without changing the total cell leakage much. To prove it, we perform Monte Carlo simulation of a cell with random variation only ($\sigma_s = 0$, $\sigma_{d2d} = 0$, and $\sigma_r = 0.052$). The simulation shows that random variation decreases the MTTF of the cell by only 0.7%.

Overall, to compute the effect of the total V_{t0} variation on the TTF$_{1\%}$ and MTTF of a cell with $P_{s0} = 50mW$, we perform Monte Carlo simulations. We generate 100,000 cells, each with a cell-wide V_{t0} value drawn from the D2D and WID systematic variation distributions, plus an additional V_{t0} component from the WID random variation distribution. For a given cell, we compute the total power dissipation of the cell and its temperature. Then, we use the equations of Section II-A to compute the MTTF and the lognormal TTF distribution of the cell. We sample the distribution to obtain a concrete failure time. We repeat this process for the 100,000 cells and compile the sample failure times into a final TTF distribution.

This final TTF distribution has a TTF$_{1\%}$ and an MTTF that are 10% lower and 2% lower, respectively, than the corresponding values without V_{t0} variation. Consequently, we see that V_{t0} variation significantly reduces the time at which 1% of the cells will have failed.

B. Whole-Processor Effects

Processors are complex series-failure systems. Moreover, heat transfer between cells renders their temperatures (and failure rates) interdependent. To determine the effect of V_{t0} variation on the TTF$_{1\%}$ of a whole processor, we use a processor model based on the 65nm Intel Core Solo. Active power

is estimated by running the *crafty* SPECint benchmark on a cycle-accurate microarchitecture simulator [13] augmented with Wattch [14]. Total dynamic power is approximately 14W, which is substantially lower than the processor's thermal design power (TDP) of 30W. Temperatures are computed using HotSpot [15], and individual cell temperatures are not allowed to go over 100°C. Unless otherwise specified, $R_{se} = 0.8K/W$ [15].

As an illustration, Figure 5 shows an example die before and after V_{t0} variation. The colors and contours indicate the normalized cell MTTF, where MTTF=1 corresponds to the MTTF of the L2 cache on the no-variation die. In this particular die, D2D and WID systematic variation have reduced the V_{t0} of the cells by one to two $\sigma_{V_{t0}}$. Before variation, the total leakage is 8.7W; after, it is 13.8W. Due to the increase in static power, the cell temperatures increase by 3–5°C, reducing cell MTTF by an average of 20%.

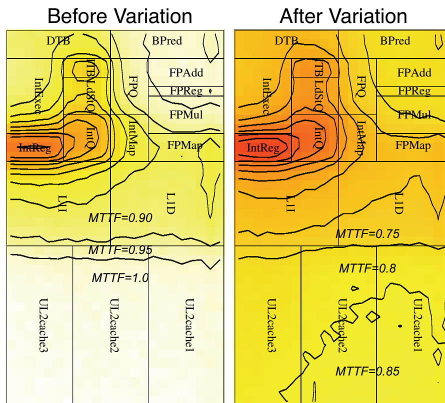

Fig. 5. Spatial distribution of cell MTTFs before and after variation for an example die. Color and contours indicate MTTF.

To compute the effect of the V_{t0} variation on the $\text{TTF}_{1\%}$ of the die, we use Monte Carlo simulations with 20,000 dies. Each die has its own V_{t0} map generated according to the D2D, WID systematic and WID random probability distributions presented in Section II-C. For each cell in a given die, we determine its temperature using the detailed model of Section II-B, generate a TTF distribution, and sample the latter to obtain a concrete failure time for the cell as in Section III-A. Then, we take the minimum of the failure times of all the cells of the die as the die's failure time. We repeat this process for all 20,000 dies and compile the dies' failure times into a final TTF distribution. From this TTF distribution, we compute the $\text{TTF}_{1\%}$.

Experimentally, we find that the resulting $\text{TTF}_{1\%}$ can be substantially lower than the $\text{TTF}_{1\%}$ before V_{t0} variation. We find that the reduction in $\text{TTF}_{1\%}$ is heavily dependent on two parameters: the heatsink thermal resistance R_{se} and the total leakage power of the die before V_{t0} variation. Let P_{leak0} be defined as the total leakage power of the die with no V_{t0} variation when the die is at a uniform 80°C. For a range of configurations (R_{se}, P_{leak0}), Table I shows the percentage reduction in $\text{TTF}_{1\%}$ relative to a no-variation die of the same configuration. The data in the table shows that a good heatsink (low R_{se}) can minimize the reliability impact of V_{t0} variation. It also shows that relatively small changes in R_{se} can have strong reliability consequences when P_{leak0} is high.

TABLE I

REDUCTION IN $TTF_{1\%}$ FOR DIFFERENT VALUES OF HEATSINK THERMAL RESISTANCE AND OF NO-VARIATION LEAKAGE POWER.

P_{leak0}	$R_{se}(K/W)$				
(W)	0.6	0.7	0.8	0.9	1.0
7.5	-2%	-3%	-4%	-4%	-9%
10.0	-2%	-6%	-11%	-19%	-34%
12.5	-8%	-12%	-25%	-50%	-62%
15.0	-14%	-37%	-61%	-67%	-64%

IV. CONCLUSIONS

Variation in threshold voltage can potentially reduce the lifetime of a 65nm processor significantly. Two important parameters that determine the processor lifetime reduction are the thermal resistance of the heatsink and the total leakage power of the processor before variation. For unfavorable values of these parameters, our models suggest that the $\text{TTF}_{1\%}$ of the processor can decrease by about 60%. These are important effects to consider as processor manufacturers carefully tune their designs for cost-effective die lifetime.

REFERENCES

[1] T. Karnik, S. Borkar, and V. De, "Probabilistic and variation-tolerant designs: Key to continued Moore's Law," in *Presentation at TAU Workshop*, Feb. 2004.

[2] SEMATECH, *Critical Reliability Challenges for the International Technology Roadmap for Semiconductors*, Mar. 2003, Technology Transfer 03024377A-TR.

[3] T. M. Mak, "How things fail?" Oct. 2006, Pres. at Univ. of Illinois.

[4] Intel Corporation, "Intel LXT971A dual-speed 3.3v 10/100 ethernet transceiver," 2001, Product brief.

[5] SEMATECH, *Semiconductor Device Reliability Failure Models*, May 2000, Technology Transfer 00053955A-XFR.

[6] Y. Zhang *et al.*, "Hotleakage: A temperature-aware model of subthreshold and gate leakage for architects," Univ. of Virginia, CS Dept., Tech. Rep. CS-2003-05, Mar. 2003.

[7] J. Srinivasan *et al.*, "The impact of technology scaling on lifetime reliability," in *Dependable Systems and Networks*, June 2004.

[8] R. Degraeve, B. Kaczer, and G. Groeseneken, "Reliability: A possible showstopper for oxide thickness scaling?" *Semiconductor Science and Technology*, vol. 15, pp. 436–444, May 2000.

[9] A. Srivastava, D. Sylvester, and D. Blaauw, *Statistical Analysis and Optimization for VLSI: Timing and Power*. Springer, June 2005.

[10] K. Katsuki, M. Kotani, K. Kobayashi, and H. Onodera, "Measurement results of within-die variations on a 90nm LUT array for speed and yield enhancement for reconfigurable devices," in *ASPDAC*, Jan. 2006.

[11] N. Cressie, *Statistics for Spatial Data*. John Wiley & Sons, Jan. 1993.

[12] P. Friedberg *et al.*, "Modeling within-die spatial correlation effects for process-design co-optimization," in *ISQED*, Mar. 2005.

[13] J. Renau, B. Fraguela, J. Tuck, W. Liu, M. Prvulovic, L. Ceze, K. Strauss, S. Sarangi, P. Sack, and P. Montesinos, "SESC Simulator," January 2005, http://sesc.sourceforge.net.

[14] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architecture-level power analysis and optimizations," in *International Symposium on Computer Architecture*, June 2000.

[15] K. Skadron *et al.*, "Temperature-aware microarchitecture," in *International Symposium on Computer Architecture*, June 2003.