

Effects of Various Applications on Relative LifeTime of Processor Cores

T. Gupta, C. Bertolini, O. Heron, N. Ventroux

CEA LIST
PC94, Gif-Sur-Yvette, F-91191, France
Email: tushar.gupta@cea.fr

T. Zimmer, F. Marc

Université Bordeaux I
351 cours de la Libération
33405 TALENCE cedex, Bordeaux, France
Email: thomas.zimmer@ims-bordeaux.fr

ABSTRACT

The lifetime of integrated chip is reducing rapidly with technology. To check if design is feasible, and to study and analyse the lifetime of processor, via studying failure mechanisms on higher level of abstraction layer, we present an interesting idea to evaluate Reliability using RTME (Real Time MTTF Evaluation) using Power Consumption and Temperature. Using the output of RTME, we are able to distinguish the effect of different benchmarks on different blocks of the processor.

KEYWORDS

Reliability, Power Consumption, Temperature, Simulation, Processor, Digital circuits.

1. INTRODUCTION

Semiconductor industry has an immense pressure for improving performance, increasing functionality, decreasing cost and reducing design and development time. For all these improvements, the solution is to minimize device feature size in nanometer scale range and further, which affects the lifetime of a chip drastically. To tackle this problem, we introduce RTME, which is a simulation tool for predicting Time to Failure (TTF) and Failure Rate (λ) of different blocks of the processor at architectural level, at a very early design stage. The objective will be to check the feasibility of the proposed design, before even synthesizing the circuit. The advantage of RTME is that it is a flexible tool, capable to compare aging behavior for different benchmarks and architecture choices, not bound to any specific simulator. RTME is believed to be hundreds of times faster than already existing tools at transistor level, but with reduced accuracy. In the rest of the document, we present how we obtain simulation results using RTME to show, which blocks of the processor get more prone to which failure mechanism and which benchmark causes faster aging in a processor. Our work shows that, Aging depends on Applications executing during lifetime of the processor, and some blocks are much more prone to failures than others.

2. PRIOR WORK AND MOTIVATION

At present, there exist many tools at transistor level to evaluate reliability. Srinivasan et al. [6] proposed an application aware architecture-level model *RAMP* to evaluate a processor's lifetime. *RAMP* introduced the methodology at higher level of abstraction, but it involves parameters to be evaluated at transistor level using *SPICE*. Some authors in the past discussed about dependence between Power consumption, Temperature and Reliability saying there is a need to relate the three together to estimate lifetime [2, 5, and 10]. D. Brooks et al [2] and K. Skadron et al [5] have shown that it is possible to estimate power consumption and temperature, at

higher level of abstraction. They validate their results in form of tools named *Watch* and *Hotspot*. There have been many studies in the last few decades showing that failures occur more and more early in the lifetime of a processor due to scaling. In our study, we have noticed that the effect of instructions (and the applications) has never been taken into account as such for a specific technology, and that motivates us to study the effect of different benchmarks on the lifetime of a processor and the cache memory.

3. FAILURE MODELS

At present time, we consider 4 important Failure models in *RTME* that affects the lifetime of processors [4, 8, 9 and 12]:

Electromigration (EM) - Due to momentum exchange between the current-carrying electrons and the host metal lattice, ions can drift in the direction of the electron current. Due to the presence of flux divergence centres, vacancies start to cluster, clusters grow into voids, and the voids can continue to grow until they block the current flow in the aluminium. Thus, the current is forced to flow through the supporting barrier layer and/or capping layer; the resultant increase in resistance leads to device failure. Since this is a mass conserving process, accumulations of the transported aluminium ions increase the mechanical stress in supporting dielectrics, and may eventually cause fractures and shorts to occur.

Hot Carriers Injection (HCI) - Hot carrier injection describes the phenomena by which carriers gain sufficient energy to be injected into the gate oxide. This occurs as carriers move along the channel of MOSFET and experience impact ionization near the drain end of the device. The damage can occur at the interface, within the oxide and/or within the sidewall spacer. Interface-state generation and charge trapping induced by this mechanism result in transistor parameter degradation, typically as switching frequency degradation rather than a "hard" functional failure.

Time-Dependent Dielectric Breakdown (TDDB) - Time-Dependent Dielectric Breakdown (TDDB) is an important failure mechanism in ULSI devices. The dielectric fails when a conductive path forms in the dielectric, shorting the anode and cathode. The two models widely used in describing TDDB are field-driven (E-model) and current-driven (1/E - model). We use the E-Model, in which because of low-field (< 10MV/cm) TDDB is due to field-enhanced thermal bond-breakage at the silicon-dielectric interface. The E-field reduces the activation energy required for thermal bond breakage and therefore TTF, inverse to reaction rate, decreases exponentially.

Negative Bias Temperature Instability (NBTI) - It is a key reliability issue that is of immediate concern in p-channel MOS devices stressed with negative gate voltages. NBTI manifests as an increase in the threshold voltage and consequent decrease in drain current and transconductance. The degradation exhibits power law dependence with time.

Failure Models are presented in terms of Time-to-Failure (TTF – a common measurement unit) to estimate reliability for semiconductor devices in theory. Table 1 lists the analytical TTF equations that model the behavior of studied failure mechanisms. It presents E-model for TDDB, follows Black's Law for Electromigration [14], Takeda model for HCI [11], and one of the phenomenological models for NBTI. Also we provide the values of constant parameters used in the simulation framework. In fact, these parameters depend on the manufacturing process and the materials and are gathered from [4, 7]. In these models, the global factor of each model is not given: these factors are technology dependent and difficult to obtain. Hence, the resulting failure rate values must be considered as given in different arbitrary units.

Table 1: Failure Models at transistor level in theory.

Name	TTF (Time-to-Failure)	Parameters
TDDB	$(e^{\gamma \cdot E}) \cdot e^{\frac{E_a}{k \cdot T}}$	γ : Field Accel. parameter: -3 Np.cm/MV, E_a : 0.7eV, E: Electric Field Applied, k: Boltzmann's constt.
EM	$(J - J_{crit})^{-a} \cdot e^{\frac{E_a}{k \cdot T}}$	J: current density req. for EM, E_a : 1.2eV, a: 2
HCI	$e^{\frac{-\gamma}{V_{DD}}} \cdot e^{\frac{E_a}{k \cdot T}}$	γ : Field Accel. parameter: 16 NpV, V_{DD} : Supply Voltage E_a : -0.2eV
NBTI	$\left(\frac{E_a}{k \cdot T} \cdot (V_G)^\beta \right)^{\frac{1}{t}}$	β : gate voltage exponent: 3.5, V_G : app. gate voltage, t: Time exponent: 0.25 E_a : 0.4eV

In *RTME*, we deal with TTF at block level, and the respective failure models are shown in Table 2. As shown, we are using some relations, such as, J (Current Density) = P/ (Vdd) and α (Switching Activity) = P/ (Vdd)² in terms of dynamic power consumption (denoted as P). In addition, we take some assumptions, since; estimation is made before physical synthesis of processor, as follows: For TDDB, we assume that half of all the transistors in each block are prone to an electrical field, at each processor clock cycle.

Table 2: Failure Models at block level modified for RTME.

Name	Failure Model (TTF_i)	Parameters
TDDB	$A/2 \cdot e^{\gamma \cdot \frac{V_{DD}}{t_{ox}}} \cdot e^{\frac{E_a}{k \cdot T}}$	A: Block area t_{ox} : gate oxide thickness
EM	$\left(\frac{P}{V_{DD}} \right)^{-a} \cdot e^{\frac{E_a}{k \cdot T}}$	J in terms of P: Instt. Power Consumption
HCI	$\alpha \cdot A/2 \cdot e^{\frac{-\gamma}{V_{DD}}} \cdot e^{\frac{E_a}{k \cdot T}}$	α : number of Transitions in terms of P
NBTI	$A/2 \cdot (V_{DD})^{\frac{\beta}{t}} \cdot e^{\frac{E_a}{k \cdot T}}$	

For EM, we assume that all transistors, in each block of the processor, contribute evenly to power consumption and dissipate the same amount of heat. For HCI, we assume that the numbers of transistors that are prone to this failure are equal to the switching activity of the block I/Os. Switching activity is the ratio between the number of bits that switch on the block I/Os and the total number of I/O bits, at a given clock cycle. For NBTI, all of PMOS transistors

are affected at each processor clock cycle, whatever the block input values. These assumptions may affect the TTF accuracy compared to real world scenario.

4. CHAIN TOOL

In Figure 1, we present RTME methodology, which is a chain of tools, to evaluate reliability at RTL abstraction level. Dynamic Power and Temperature Simulators we use are state-of-the-art tools, providing values to variables, such as J, α and Temperature (T).

Dynamic Power traces are obtained using the tool named Wattch [2]. We chose Wattch, after considering various other tools, such as Sim-Panalyzer, since Wattch is an architectural level tool that has been proposed to analyze dynamic power with respect to simulation performance tradeoffs with a reasonable level of accuracy when compared to lower level estimation approaches. Authors claim that, it maintains accuracy within 10% of their estimates as verified using industry tools on leading-edge designs. It estimates the Dynamic power Consumption using different power models such as, $P = C * V_{dd}^2 * \alpha * f$, where C is the equivalent block capacitance, α is number of transitions and f is operating frequency. C, Vdd and f depend on process technology.

Wattch is an extension of SimpleScalar simulator [1]. The SimpleScalar tool set is used to simulate behavior of each block of the processor based on a MIPS instruction set. It enables the comparison of benchmark performance vs. different micro-architecture choices. Wattch includes various hardware counters in SimpleScalar to obtain switching activity. To estimate capacitance, Wattch uses various block models based on circuit and transistor sizing from provided technology node.

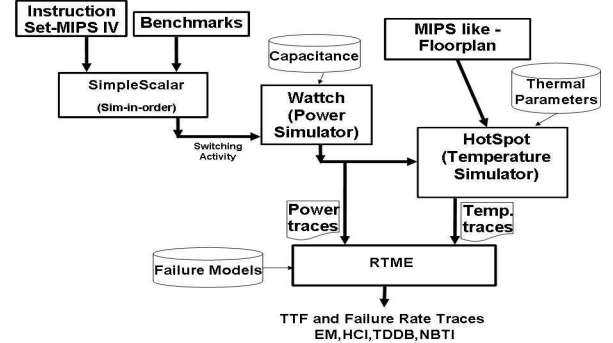


Figure 1: RTME Methodology.

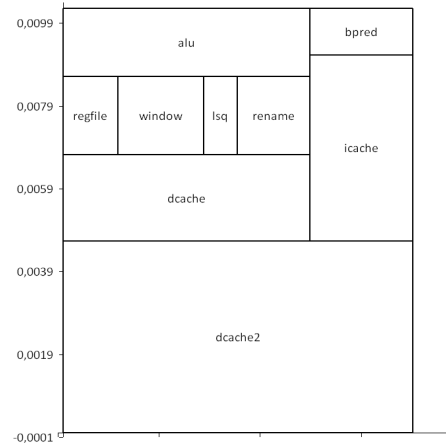


Figure 2: Floorplan

Temperature traces are obtained using the tool named *HotSpot* [5]. The HotSpot thermal model of each block is an electrical equivalent RC model where current is equivalent to power and node voltage is to temperature. Using Floorplan (based on SimpleScalar architecture) as shown in Figure 2, of the chip and power traces, HotSpot builds an equivalent RC electrical circuit, accounting for vertical and lateral heat transfers. According to the authors, if we provide total Power consumption at given time instant, and Thermal RC equivalent of a circuit, it is possible to estimate temperature change for the previous time interval. HotSpot also assumes a typical thermal packaging composition formed by a thermal spreader inserted between the chip substrate and the packaging. A heat sink is placed on the top of the packaging.

In RTME, we implement the respective failure models are shown in Table 2. Hence, we are able to show the variations in lifetime of various blocks of the processor core using above discussed Power and Temperature Traces and other technology related. At each simulation step, we compute the current failure rate (λ) of each block which is actually the inverse of current TTF, for each block of the processor and each failure, and can be expressed as: $\lambda(i,j) = 1/TTF(i,j)$ where 'i' is the current simulation step and 'j' is the block reference. RTME is a flexible tool allowing to compare processor aging behavior for different benchmarks and architecture choices, For that, RTME computes the Cumulative Failure Rate (CFR) of each block and for each failure mechanism, which is:

$$CFR_x(n, j) = \sum_{i=1}^n \lambda_x(i, j) * t_i$$

Where 'i' is the time step of duration 't_i', 'n' is the simulation length, 'j' is the block reference and 'x' is the failure mechanism reference.

As mentioned in Section 2, the TTF accuracy of each block is affected by the assumptions made for estimation at RTL abstraction layer. Other factors that affect RTME accuracy are the different technology parameters and the accuracy of other tools discussed above. Since we use the state-of-the-art tools with their own level of accuracy, the error is estimated in different ways and in addition failure models have different dependence with various parameters. Hence to estimate the accuracy of RTME is difficult at present. But we only need to study the relative results to compare the benchmarks and their effect in different blocks, for a specific failure mechanism, so we do not derive error estimation yet.

We assume Floorplan with Heat sink and Heat spreader with default specs provided with Hotspot, and initial Power (Leakage) and Temperature to be 0.7W and 42°C respectively, for all failures. We are working with 180nm Technology at present, which is a parameter for RTME, and flexible to change in the future, by changing the Floorplan, HotSpot configuration file and libraries for Wattch.

5. RESULTS

In this section, we present the simulated aging results obtained from 9 benchmarks (MiBench [3]) executed on a given microarchitecture. The microarchitecture blocks and the distribution of block area are illustrated on Figure 2. SimpleScalar simulates a 5 stage pipeline RISC processor with superscalar capabilities [1]. Dcache and Dcache2 are respectively the first and second level of data cache memories. The former is rarely used by the different benchmarks compared to the later. LSQ is Load/Store Queue unit handles memory synchronization/communication and contains all

loads and stores in program order. The ALU is the arithmetic logic unit composed of scalar operators. Regfile is the register file composed of 32 64-bit registers. SimpleScalar allows configuring the size and behavior of the different pipeline stages.

We use various graphs and diagrams to show relation between Power consumption, Temperature, and Reliability. In Figures 3 and 4 graphs shows average power consumption and temperature for the complete execution of each benchmark, for each pipeline block. The power and area distribution are the elements to evaluate Temperature. Whatever the benchmark, we can observe that the ALU consume most of the total dynamic power. Actually, the major instruction class from these benchmarks is the computing instructions. In addition, we can remark that the benchmark CRC32 causes the highest average heat dissipation, whatever the block.

Figure 5 presents the Cumulative Failure Rate for each benchmark and each pipeline block. In each graph, the values correspond to the CFR of only one failure mechanism, thus explaining the plot of four graphs. NBTI and Tddb have similar behavior and are very much depending on changes in temperature. EM has more effect of Power consumption, and HCI has involvement of both Power consumption and Temperature. The results show that EM in "ALU" block, Tddb (and NBTI) in "BPred" and "ALU" block of the processor are more effective and for each failure the "Bitcount" is the benchmark, causing faster aging in the processor compared to others.

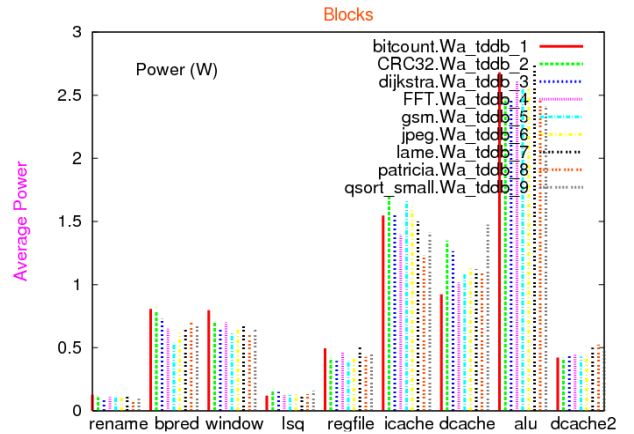


Figure 3: Average power consumption vs. Different pipeline blocks and benchmarks

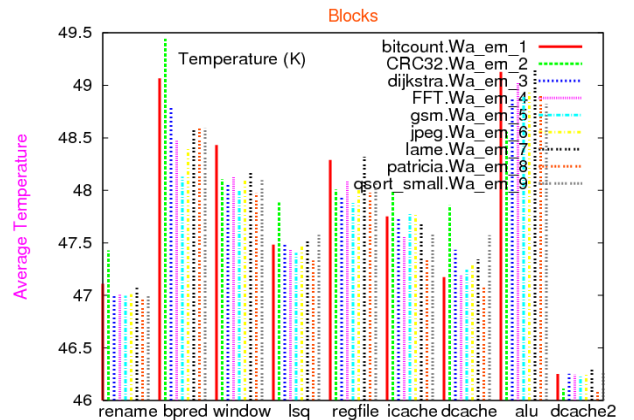


Figure 4: Average power consumption vs. Different pipeline blocks and benchmarks.

6. CONCLUSION

In this paper, we presented a flexible simulation tool chain for estimating the reliability of a processor core, at RTL abstraction level. For a given microarchitecture, we show that the aging of a processor depends on the benchmark profile. In the present time, we didn't actually use a manufacturer technology library so far, for now we can compare reliability results for different benchmarks, and observe which application is affecting the processor's reliability most. Due to lack of up-to-date industrial fabrication and reliability data in public-domain, we are not able to validate the results, still there is need to make refinements in models and tools, which may lead to variations in absolute values of the results, but this, will not vary the relative nature of the results. So, we can say that using proper technology libraries in future, we can tell which failure mechanism will be more effective.

REFERENCES

- [1] T. Austin, E. Larson, and D. Ernst. SimpleScalar: an infrastructure for computer system modeling, Feb 2002.
- [2] D. Brooks, V. Tiwari, and M. Martonosi. Wattch: a framework for architectural-level power analysis and optimizations. Computer Architecture, 2000. Proceedings of the 27th International Symposium.
- [3] M. Guthaus, J. Ringenberg, D. Ernst, T. Austin, T. Mudge, and R. Brown. MiBench: A free, commercially representative embedded benchmark suite. Workload Characterization. WWC-4. 2001 IEEE International Workshop.
- [4] JEDEC Publication. Failure Mechanisms and Models for Semiconductor Devices, March 2009.
- [5] K. Skadron, M. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan. Temperature-aware microarchitecture. Computer Architecture, 2003. Proceedings. 30th Annual International Symposium.
- [6] J. Srinivasan, S. V. Adve, P. Bose, and J. A. Rivers. The case for lifetime reliability-aware microprocessors. In ISCA, 2004.
- [7] M. White and J. B. Bernstein. Microelectronics reliability, physics-of-failure based modeling and lifetime evaluation, 2008.
- [8] Richard Blish, Noel Durrant. Semiconductor Device Reliability Failure Models, May 2000.
- [9] JEDEC Publication, Methods for Calculating Failure Rates in Units of FITs, July 2001.
- [10] Ayse K. Coskun, Tajana Simunic Rosing and Keith Whisnant. Temperature Aware Task Scheduling in MPSoCs. In Proceedings of Design Automation and Test in Europe (DATE), pp. 1659-1664, 2007.
- [11] E. Takeda, C. Y. Yang, and A. Miura-Hamada, Hot-Carrier Effects in MOS Devices, chapter 2, pp. 49-58. Academic Press, 1995.
- [12] Critical Reliability Challenges for the International Technology Roadmap for Semiconductors (ITRS), International SEMATECH, 2003.
- [13] EECS UMICH. "Sim-Panalyzer," <http://www.eecs.umich.edu/~panalyzer/>
- [14] J. R. Lloyd, "Reliability modeling for Electromigration failure," Quality and Reliability Engineering International, vol. 10, pp. 303-308, 1994.

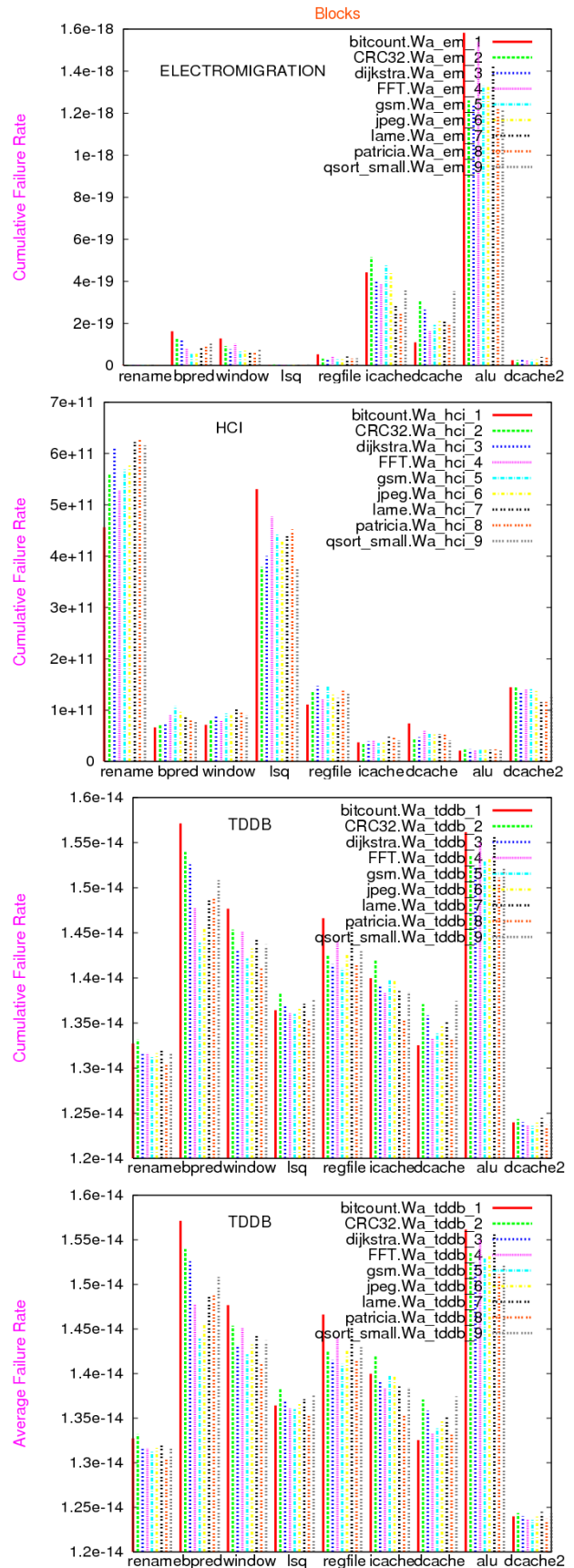


Figure 5: Different Cumulative Failure Rate.