Interconnect Lifetime Prediction for Temperature-Aware Design

UNIV. OF VIRGINIA DEPT. OF COMPUTER SCIENCE TECH. REPORT CS-2003-21
NOVEMBER 2003

Zhijian Lu†, Mircea Stan†, John Lach†, Kevin Skadron‡
Departments of †Electrical and Computer Engineering and ‡Computer Science, University of Virginia
Charlottesville, VA 22904

Abstract

Thermal effects are becoming a limiting factor in high-performance circuit design due to the strong temperature-dependence of leakage power, circuit performance, IC package cost and reliability. Temperature-aware design tries to maximize performance under a given thermal envelope through various static and dynamic approaches. While existing interconnect reliability models assume a constant temperature, this paper presents a technique for probabilistically estimating interconnect lifetime for any time-varying temperature profile. With this formulation, interconnect lifetime can be modeled as a resource that is consumed over time, with the rate of consumption being a function of temperature. As a result, designers may be more aggressive in the temperature profiles that are allowed on a chip instead of using static worst-case assumptions. For example, performance (hence power and temperature) may be increased beyond what is allowed by worst-case restrictions for short periods as long as the increase is compensated for later by lower activity. With this model, temperature-aware designs will achieve higher overall performance while satisfying lifetime requirements. *This report is superseded by TR CS-2004-08.*

I. Introduction

Due to increasing system complexities and clock frequencies, temperature has become a major concern in integrated circuit design. Higher temperatures not only degrade system performance, raise packaging costs, and increase leakage power, but they also affect system reliability [1] by increasing interconnect electromigration (EM) and gate oxide failure rates.

As a result, the field of temperature-aware design has emerged, which seeks to maximize system performance under lifetime constraints. Considering system lifetime as a resource that is consumed over time as a function of temperature, dynamic thermal management (DTM) techniques [10], [11], [6] are being developed to best manage this consumption. While the dynamic temperature profile of a system is workload-dependent [10], [11], several efficient and accurate techniques have been proposed to simulate transient chip-wide temperature distribution [10], [12], [3], providing design-time knowledge of the thermal behavior of different design alternatives.

To best evaluate these techniques and explore the design space under system lifetime constraints, a reliability analysis method based on thermal information is needed. Black [2] proposed a semi-empirical temperature-dependent model for EM failures:

$$T_f = \frac{A}{j^n} exp\left(\frac{Q}{kT}\right) \tag{1}$$

where T_f is the time to failure, A is a constant based on the interconnect geometry and material, j is the current density, Q is the activation energy, and kT is the thermal energy. Black's model is widely used in thermal reliability analysis and design. For example, Banerjee *et al.* [1] derived a self-consistent allowable current density upper bound for achieving a reliability goal by taking into account interconnect self-heating effects using Black's model.

However, Black's model assumes a constant temperature. Thus, a worst-case temperature profile is usually used when applying this model, resulting in pessimistic estimations and more a restrictive design space. Lall *et al.* [9] showed that many failure mechanisms are not due to high steady-state temperature. Rather, temperature gradients, temperature cycle magnitude and temperature rates of change play a more prominent role.

In this paper, we present a simple formula to probabilistically estimate interconnect lifetime based on EM failures with a time-varying temperature distribution. (The effects of variable temperatures on gate oxide will be the subject of future work.) When incorporated into analysis tools, temperature-driven design can more accurately evaluate DTM techniques to maximize system performance under lifetime constraints. In addition, designers may more aggressively explore the design space, as previous static worst-case temperature restrictions may be violated as long as the violations are compensated for later in the system?s lifetime.

This paper is organized as follows. Section II introduces a constant temperature analytic model for electromigration. In Section III, we apply this model to address time-varying temperature, and we derive a formula to estimate interconnect lifetime, which we analyze in Section IV. Finally, we summarize the paper in Section V.

II. Analytic model for electromigration with constant temperature

Clement [5] provides a review of 1-D analytic EM models. In this paper, we only discuss the EM-induced stress build-up model [4], [7], which fits well with reality. Electromigration is the process of self-diffusion due to the momentum exchange between electrons and atoms. The dislocation of atoms causes stress build-up according to the following equation [4], [7].

$$\frac{\partial \sigma}{\partial t} - D_a \left(\frac{B\Omega}{kTl^2\varepsilon}\right) \frac{\partial}{\partial x}\left(\frac{\partial \sigma}{\partial x} - \frac{qlE}{\Omega}\right) = 0 \tag{2}$$

where $\sigma(x,t)$ is the stress function; D_a is the diffusivity of atoms, a function of temperature; B is the appropriate elastic modulus, depending on the properties of the metal and the surrounding material and the line aspect ratio; Ω is the atom volume; ε is the ratio of the line cross-sectional area to the area of the diffusion path; l is the characteristic length of the metal line; q is the effective charge; and E is the applied electric field, which is equal to ρj, the product of resistivity and current density. The term $\frac{qlE}{\Omega}$ corresponds to the atom flux due to the electric field, while $\frac{\partial \sigma}{\partial x}$ corresponds to a backflow flux created by the stress gradient to counter-balance the EM flux. This equation assumes that the temperature is uniform across the interconnect line. If we let $\beta(T) = D_a\left(\frac{B\Omega}{kTl^2\varepsilon}\right)$ and $\alpha(j) = \frac{qlE}{\Omega}$, we will obtain the following simplified version:

$$\frac{\partial \sigma}{\partial t} - \beta(T)\frac{\partial}{\partial x}\left(\frac{\partial \sigma}{\partial x} - \alpha(j)\right) = 0 \tag{3}$$

Some example values for β under different temperatures are calculated using the data found in [4], [7], and shown in Table I. The meaning of $\beta(T)$ will be discussed later.

Temperature(K)	350	370	400	420	440	450
$\beta(T)$	0.0029e-8	0.0080e-8	0.0302e-8	0.0659e-8	0.1336e-8	0.1857e-8

TABLE I

Some values of β at different temperature. Interconnect length of 100μm is used.

Clement [4] investigated the effect of current density on stress build-up using Equation (3), assuming that temperature is unchanged (i.e. $\beta(T) = constant$), for several different boundary conditions. He found that the time to failure derived from this analytic model had exactly the same form as Black's equation (1). He also found that the exponential component in Black's equation was due to atom diffusivity's (D_a's) dependency on temperature by the well-known Arrhenius equation:

$$D_a = D_{ao} exp\left(\frac{-Q}{kT}\right) \tag{4}$$

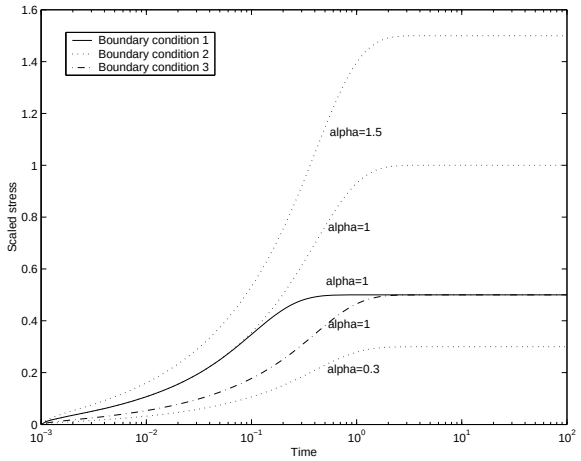

Fig. 1. EM stress build-up for different boundary conditions and α values. All processes have $\beta = 1$.

Applying the parabolic maximum principles [8] to the above partial differential equation (3), we know that at any time t, the maximum stress along a metal line can be found at the boundaries of the interconnect line. Figure 1 shows the numerical solutions for equation (3) at one end of the line (i.e. $x = 0$) for different boundary conditions and α values, with all having $\beta = 1$. The three boundary conditions shown here are similar to those discussed in [4] for finite length interconnect lines. It clearly indicates that both boundary conditions and current density (i.e.α) affect the stress build-up rate. Also seen from the figure is that the stress build-up saturates at a certain point. This is because, in saturation, the atom flux caused by electromigration is completely counterbalanced by the flux created by the stress gradient along the metal line. In our discussion, we assume that the critical stress for EM failure is below these saturated values.

III. Electromigration under time-varying temperature

If both current density and temperature change over the time, Equation (3) can be still used to model the EM stress build-up in the interconnect line. Clement [4] found that in the case where temperature is kept constant, the average

current density should be used in Black's equation. In the case where temperature varies over time, we can assume that the average current density is used (i.e. α is a constant), while the temperature is a function of time (i.e. $\beta(t)$).

Since temperature may be an arbitrary function of time, it is impossible to directly solve the stress build-up equation. Instead, we solve the problem indirectly based on the following theorem.

Theorem 1: Consider the stress build-up equation (3) with constant values for β and α. Let $\sigma_1(x,t)$ be the solution for the equation with $\beta = \beta_1$ under certain initial and boundary conditions, $\sigma_2(x,t)$ be the solution with $\beta = \beta_2$ for the same initial and boundary conditions. If the solutions for Equation (3) are unique for those initial and boundary conditions, we have

$$\sigma_2(x,t) = \sigma_1(x, \left(\frac{\beta_2}{\beta_1}\right) t) \tag{5}$$

This theorem can be easily proved by plugging in Equation (3) with $\sigma_2(x,t)$ (Equation (5)) and β_2 respectively, and checking the satisfiability for the initial and boundary conditions.

Theorem (1) tells us that the stress build-up processes in the interconnect are independent of the value of β in Equation (3). The value of β only determines the build-up speed of the process. For example, at time $\left(\frac{\beta_2}{\beta_1}\right) t$, stress build-up in process 1 (i.e. $\beta = \beta_1$) *sees* the stress build-up in process 2 (i.e. $\beta = \beta_2$) at time t. If we are only concerned with the stress build-up at a fixed point on the line (e.g., at one end of the line, which always has the maximum stress at any time), we have the following corollary.

Corollary 1.1: For all stress build-up processes on the same metal line specified by Equation (3), with the only difference in the value of β, the stress build-up at a fixed point is only decided by the product of time and the value of β:

$$\sigma = f(\beta_i t_i) \tag{6}$$

where f is a function independent of temperature.

The corollary can be proven using Equation (5) with x fixed. Let σ_i denote the stress build-up at a specific point of the interconnect for the process with β_i. Corollary (1.1) shows that $\sigma_1(t_1) = \sigma_2(t_2) = \sigma_3(t_3) = \ldots$, as long as $\beta_1 t_1 = \beta_2 t_2 = \beta_3 t_3 = \ldots$.

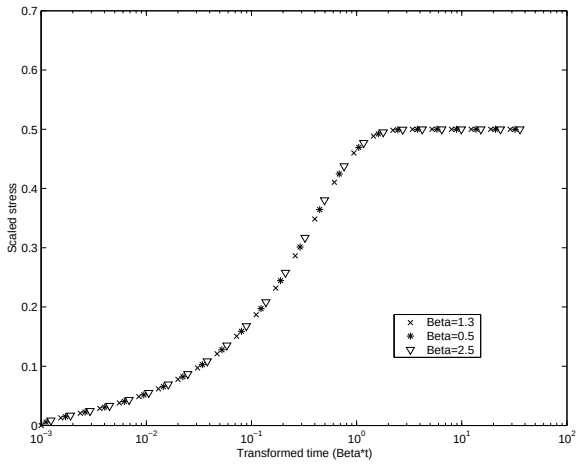

Fig. 2. EM stress build-up at one end of the interconnect for different β values as a function of βt.

Figure 2 illustrates Equation (6) by showing the stress build-up at one end of the metal line for several different values of β, with the same initial and boundary conditions. As expected from Equation (6), the stress is a unique function of βt, in spite of the different values of β.

Consider that temperature varies over time. We can divide time into segments, such that temperature is constant within each time segment. In other words, β in Equation (3) is a segment-wise function, described as:

$$\beta(t) = \begin{cases} \beta_1, & t \in [0, \Delta t_1] \\ \beta_2, & t \in (\Delta t_1, \Delta t_1 + \Delta t_2] \\ \cdots & \\ \beta_i, & t \in \left(\sum_{k=1}^{i-1} \Delta t_k, \sum_{k=1}^{i} \Delta t_k\right] \\ \cdots & \end{cases}$$

Since we are only concerned with the maximum stress build-up, let $\sigma_i(t)$ denote the distinct stress build-up at the end of a metal line with a constant value of β_i, and let $\sigma_(t)$ be the stress build-up with the time-varying function $\beta(t)$. During

the first time segment, the stress builds up the same as process $\sigma_1(t)$. At the end of this time segment, which is denoted by t_1 (i.e., $t_1 = \Delta t_1$), we have $\sigma(t_1) = \sigma_1(t_1) = f(\beta_1 \Delta t_1) = \sigma_2(\frac{\beta_1}{\beta_2}\Delta t_1)$, from Corollary (1.1). Also at time t_1, according to Theorem (1), the stress distribution along the metal line is the same as that for the process $\sigma_2(x,t)$ at time $\left(\frac{\beta_1}{\beta_2}\right)\Delta t_1$. Since during the second time segment, $\beta(t) = \beta_2$, the stress build-up (i.e.$(\Delta t_1, \Delta t_1 + \Delta t_2]$) is *exactly* the same as that in process $\sigma_2(t)$ during the time interval $\left(\left(\frac{\beta_1}{\beta_2}\right)\Delta t_1, \left(\frac{\beta_1}{\beta_2}\right)\Delta t_1 + \Delta t_2\right]$. So at the end of the second time segment (i.e. $t = t_2 = \Delta t_1 + \Delta t_2$), we have

$\sigma(t_2) = \sigma_2\left(\left(\frac{\beta_1}{\beta_2}\right)\Delta t_1 + \Delta t_2\right) = f\left(\beta_2\left(\frac{\beta_1}{\beta_2}\Delta t_1 + \Delta t_2\right)\right) = f\left(\beta_1 \Delta t_1 + \beta_2 \Delta t_2\right)$, again due to Equation (6). Similar analysis can be applied to other time segments. It follows then that at the end the ith time segment, the stress is specified as:

$$\sigma(t_i) = f\left(\sum_{k=1}^{i} \beta_k \Delta t_k\right) \tag{7}$$

where f is defined in Corollary (1.1).

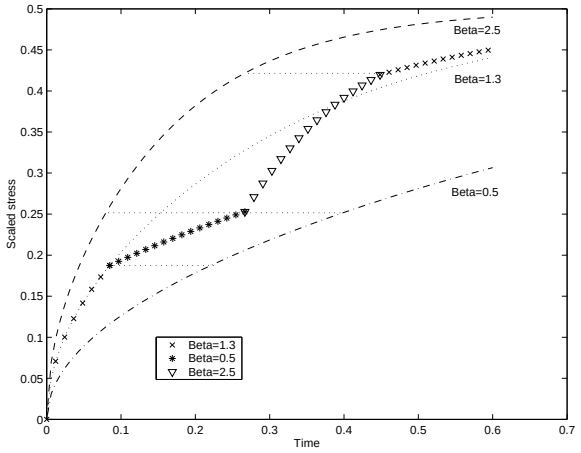

Fig. 3. EM stress build-up at one end of the interconnect with time-varying temperature. Also shown are the build-ups with invariant temperatures.

As an example, a numerical solution for Equation (3) with three different temperatures (i.e. $\beta_1 = 1.3, \beta_2 = 0.5, \beta_3 = 2.5$) over time is plotted in Figure 3. Also shown in the figure are the build-ups with time-invariant temperatures. As clearly indicated in the figure, the stress build-up process in each time segment is a shifted version of the corresponding process with time-invariant temperature. This figure is consistent with the above analysis.

As $\Delta t_i \rightarrow dt$, $\beta_i \rightarrow \beta(T(t))$, we obtain the integral version for the stress build-up function:

$$\sigma(t) = f\left(\int_0^t \beta(T(t))dt\right) \tag{8}$$

If we assume that the stress build-up reaches a certain threshold (σ_{th}) when an EM failure occurs, we have:

$$\int_0^{t_{failure}} \beta(T(t))dt = \varphi_{th} \tag{9}$$

where φ_{th} is decided by the critical stress (i.e. $\varphi_{th} = f^{-1}(\sigma_{th})$). If an average value of $\beta(t)$ exists, we obtain a closed form for the time to failure:

$$t_{failure} = \frac{\varphi_{th}}{E(\beta(T(t)))} \tag{10}$$

where $E(\beta(t))$ is the expected value for $\beta(t)$, and $\beta(t)$ is a function of temperature having the form:

$$\beta(T(t)) = A\left(\frac{exp\left(-\frac{Q}{kT(t)}\right)}{kT(t)}\right) \tag{11}$$

where A is a constant.

One way to interpret Equation (9) is to consider interconnect time to failure (i.e. interconnect lifetime) as an available resource, which is consumed by the system over time. Then the $\beta(t)$ function is the rate at which the resource is consumed, which provides the probabilistic analysis for temperature-aware design.

Let $MTF(T)$ be the time to failure with a constant temperature T. We have $\beta(T) = \frac{\varphi_{th}}{MTF(T)}$ by Equation (10). Substitute this relation in Equation (10) again and consider the time-varying temperature, and we obtain an alternative form for Equation (10):

$$t_{failure} = \frac{1}{E(1/MTF(T))} \tag{12}$$

Equation (12) can be used to derive the absolute time to failure provided that we know the time to failure for different constant temperatures (e.g. data from experiments).

IV. Reliability analysis under time-varying temperature

In this section, we apply the above derived formula to evaluate the EM reliability under time-varying temperature. It is interesting to compare the reliability of constant temperature with that of fluctuating temperature. By calculating the second derivative of $\beta(T)$ as a function of temperature (i.e. Equation (11)), it can be verified that $\beta(T)$ is a convex function within the operational temperatures. By applying Jensen's inequality, we have

$$E(\beta(T)) \geq \beta(E(T))$$

which, according to Equation (10), leads to our first observation: constant temperature is always better in terms of EM reliability than oscillating around that temperature (with the average temperature the same as the constant temperature).

Using Equation (10) or (12), we can easily compare the EM reliability for any two temporal temperature profiles by:

$$\frac{MTF(T_1(t))}{MTF(T_2(t))} = \frac{E(\beta(T_2(t)))}{E(\beta(T_1(t)))} \tag{13}$$

where $MTF(T_1(t))$ is the time to failure under time-varying temperature profile $T_1(t)$.

Due to the duality between heat transfer and electrical phenomena, temperature varies exponentially. Heomoo *et al.* [6] investigated the effect of activity migration on thermal and power management, which explicitly created cyclic temperature waves. We use an exponential function similar to that in Heomoo *et al.*'s work to model the temperature fluctuation as follows.

$$T(t) = \begin{cases} A_1 exp\left(-\frac{t-mP}{\tau}\right) + A_2, mP < t \leq mP + rP \\ A_3 exp\left(-\frac{t-(m+r)P}{\tau}\right) + A_4, (m+r)P < t \leq (m+1)P \end{cases}$$

where $m = 0, 1, 2, \ldots$; A_1, A_2, A_3 and A_4 are constants; P is the cycle period; τ is the thermal constant; and r is the duty factor. Figure 4 shows an example waveform for temperature modeled by the above exponential function.

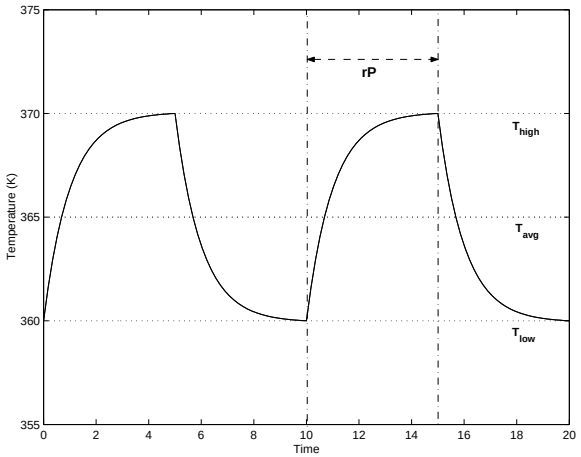

Fig. 4. Temperature wave modeled by the exponential function.

Using Equation (10), we can evaluate the EM reliability with the temperature function described above and compare it with that of a constant temperature. Figure 5 presents the results for temperatures varying with different high and low temperatures, with 50% duty cycles. In this figure, MTF_1 is the time to failure with a cyclic waveform, while MTF_2 is the time to failure with a constant temperature equal to the average temperature. Agreeing with the discussion above,

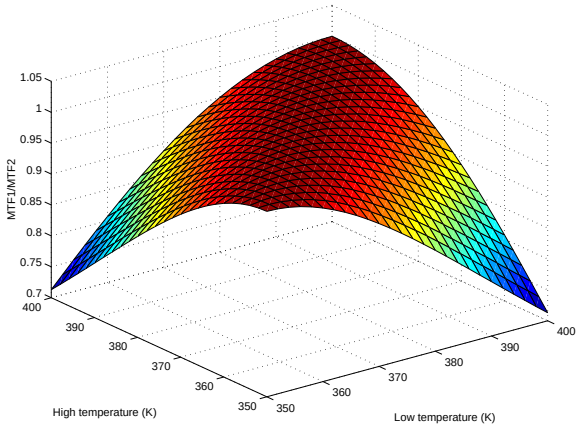

Fig. 5. EM reliability for cyclic temperature waveforms of different high and low temperatures, with duty factor $r = 0.5$. MTF_1 is the time to failure with a cyclic temperature, and MTF_2 is the time to failure with a constant temperature equal to the average temperature.

Figure 5 shows that constant temperature is always more reliable. When the temperature changes within a small range, as we may expect, the time to failure is almost equal to that of a constant temperature. The degradation of reliability is mostly decided by the fluctuation amplitude of the temperature wave. Larger cycle magnitude degrades the reliability faster.

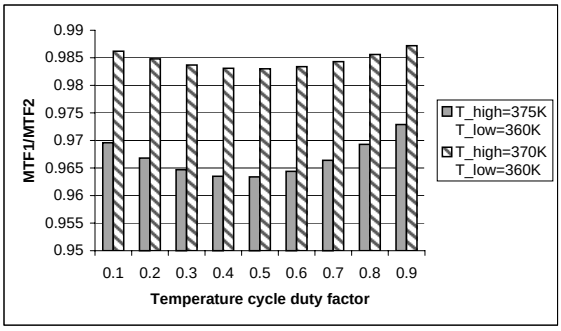

Fig. 6. EM reliability for cyclic temperatures with different duty factors r. MTF_1 is the time to failure with cyclic temperatures, and MTF_2 is the time to failure with constant temperatures equal to the average temperature.

Another interesting question is how the temperature duty factor affects reliability. Figure 6 plots the reliability comparison for various duty factors. Although half duty factor is always the worst case for reliability, the difference due to duty factor seems to be small. On the other hand, increasing the cycle magnitude by 5 degrees will reduce the relative reliability by 2% in the case shown in the figure.

As discussed above, $\beta(t)$ provides the designer an opportunity to control how the circuit lifetime is consumed. Using existing techniques, a worst-case temperature is specified to meet a certain lifetime goal. However, our model shows it is possible to overshoot the temperature threshold and compensate the lifetime "consumption" with lower temperatures later. Table II gives the reliability-equivalent constant temperature for different temperature cycle magnitudes. For example, if the temperature threshold is $361.5K$, we can achieve the same lifetime with an oscillating temperature of up to $370K$ for half of the circuit's lifetime. This enables temperature-aware designers to more aggressively explore the design space, resulting in higher performance systems that still meet specified lifetime constraints.

High Temperature(K)	360	365	370	375	380
Temperature Cycle Magnitude(K)	10	15	20	25	30
Reliability Equivalent Temperature(K)	355.5	358.2	361.5	364.8	368

TABLE II

EM reliability equivalent temperatures. A 50% duty cycle is used for the temperature cycles.

V. Conclusions and future work

Due to recent developments in design automation, integrated circuit designers can accurately model system runtime characteristics at design time to better explore the design space and evaluate design options. This paper presented an interconnect reliability model for EM failures with time-varying temperature, which will not only increase the accuracy of probabilistic reliability estimates but will enable designers to more aggressively explore the design space. For example, existing constant-temperature models require designers to observe a static worst-case temperature limit, but the model presented here enables temperature-aware designers to evaluate DTM techniques that may violate worst-case limits but still meet lifetime constraints by compensating with prior or future low temperature periods. Therefore, systems may achieve overall performance improvements under lifetime constraints. Future work will include analyzing interconnect EM reliability with a spatially non-uniform temperature distribution and the effect of time-varying temperature on gate oxide, the other major temperature-related source of IC failure.

References

[1] K. Banerjee and A. Mehrotra. Global (interconnect) warning. *IEEE Circuits and Device Magazine*, page 16, September 2001.

[2] J. R. Black. Mass transport of aluminum by momentum exchange with conducting electrons. In *IEEE Int. Rel. Phys. Symp.*, pages 148–159, 1967.

[3] Y.-K. Cheng and S.-M. Kang. A temperature-aware simulation environment for reliable ulsi chip design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 19(10):1211–20, Octorber 2000.

[4] J. Clement. Reliability analysis for encapsulated interconnect lines under dc and pulsed dc current using a continuum electromigration transport model. *J. Appl. Phys.*, 82(12):5991, 1997.

[5] J. Clement. Electromigration modeling for integrated circuit interconnect reliability analysis. *IEEE Transactions on Device and Materials Reliability*, 1(1):33–42, March 2001.

[6] S. Heo, Kenneth, Barr, and K. Asanović. Reducing power density through activity migration. In *Proceedings of the 2003 International Symposium on Low Power Electronics and Design*, pages 217–222, Seoul, Korea, 2003.

[7] M. A. Korhonen and P. Bøgesen. Stress evolution due to electromigration in confined metal lines. *J. Appl. Phys.*, 73(8):3790, April 1993.

[8] R. C. McOwen. *Partial differential equations: methods and applications*. Prentice-Hall, 1995.

[9] M. P. P. Lall and E. Hakim. *Influence of Temperature on Microelectronics and System Reliability*. CRC Press, 1997.

[10] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankanarayanan, and D. Tarjan. Temperature-ware microarchitecture. In *Proceedings of the 30th International Symposium on Computer Architecture*, pages 2–13, San Diego, CA, 2003.

[11] J. Srinivasan and S. V. Adve. Predictive dynamic thermal management for multimedia applications. In *Proceedings of the 17th Annual ACM International Conference on Supercomputing (ICS03)*, pages 109–120, June 2003.

[12] T.-Y. Wang and C. C.-P. Chen. 3-d thermal-adi: A linear-time chip level transient thermal simulator. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 21(12):1434–45, December 2002.