Analysis of Non-Uniform Temperature-Dependent Interconnect Performance in High Performance ICs

Amir H. Ajami[1], Kaustav Banerjee[2], Massoud Pedram[1], Lukas P.P.P. van Ginneken[3]
[1]Dept. of EE-Systems, Univ. of Southern California, Los Angeles, CA 90089, {aajami, massoud}@zugros.usc.edu
[2]Center of Integrated Systems, Stanford University, Stanford, CA 94305, kaustav@cis.stanford.edu
[3]Magma Design Automation Inc., Cupertino, CA 95014, lukas@magma-da.com

Abstract

Non-uniform temperature profiles along global interconnect lines in high-performance ICs can significantly impact the performance of these lines. This paper presents a detailed analysis and modeling of the interconnect performance degradation due to non-uniform temperature profiles that exist along their lengths, which in turn arise due to the thermal gradients in the underlying substrate. A non-uniform temperature-dependent distributed RC interconnect delay model is proposed for the first time. The model has been applied to a wide variety of interconnect layouts and temperature distributions to quantify the impact on signal integrity issues including clock skew fluctuations.

1 Introduction

The increasing demand for more complex VLSI circuits with higher performance is leading to higher power dissipation and increased thermal problems. Thermal issues are rapidly becoming one of the most challenging problems in high-performance chip design due to ever-increasing device count and clock speed [1]. Thermal management is essential to the development of future generations of microprocessors, integrated network processors, and systems-on-a-chip. At the circuit level, temperature variations in the substrate and interconnect lines have important implications for circuit performance and reliability [2],[3],[4].

As VLSI technology continues to scale, aggressive downscaling of interconnect dimensions causes an increase in the current density. This increase has a notable impact on the interconnect reliability mainly due to the electromigration (EM) phenomenon, which is the directional migration of interconnect metal atoms caused by the flow of electrons. Much work has been done to calculate the EM lifetime reliability of metal interconnects based on Black's equation [5],[6],[7]. Furthermore, increased current density in interconnects causes increased self-heating (or Joule heating). This self-heating effect results in temperature rise in interconnects, which exponentially reduces the interconnect EM time-to-failure [3]. This fact reveals the importance of temperature effects on the circuit reliability, which in turn calls for a detailed and accurate thermal analysis of the VLSI chips.

Although extensive research has been performed to determine the chip temperature profile and predict the effect of temperature on the interconnect reliability, few efforts have focused on analyzing the temperature effects on interconnect performance. It is well known that the resistance of an interconnect increases linearly with the temperature. In high performance ICs the chip temperature can rise up to 120 ºC [6],[8]. Additionally it has been shown that temperature variations can cause a significant change in the interconnect resistance and thereby alter the propagation delay of different paths and, in some extreme cases, cause timing violations [9].

DAC'01, June 18-22, 2001, Las Vega, Nevada, USA.

Furthermore, it has been shown that temperature gradients on the silicon substrate can occur due to different activities and/or different sleep modes of various functional blocks in high-performance microprocessor chips [10]. Dynamic power management (DPM) [11] and functional-block clock gating can be major sources of such thermal gradients over the substrate. Based on the cell power consumption map over the substrate, some researchers have provided techniques to derive the temperature profile along the substrate surface [12]. The existence of such thermal gradients along the substrate can introduce non-uniform temperature profiles along the lengths of the long global interconnects. Also, as scaling continues, the top metal layers are getting closer to the substrate which will further increase the impact of substrate thermal gradients on the interconnect temperature profiles [8]. The presence of non-uniform temperature distributions along global wires affects solutions to the clock skew control, wire sizing, layer assignment, and buffer insertion problems. Hence, it is important to analyze and quantify the impact of these thermal non-uniformities on the interconnect performance.

This paper presents the modeling and analysis of the non-uniform interconnect thermal profile and its effects on the signal performance. Section 2 introduces the effects of non-uniform interconnect thermal profile on the Elmore delay. A distributed *RC* interconnect delay model as a function of the line temperature is proposed, and a design methodology for improving the performance in the presence of non-uniform interconnect thermal profiles is presented. Section 3 discusses the background for interconnect temperature calculation and provides a systematic way of calculating the temperature along the length of an interconnect in the presence of substrate thermal gradients. The actual boundary conditions and interconnect-via/contact arrangements inside the chip are used to obtain the thermal profiles along the interconnects. Section 4 illustrates the effects of the non-uniform interconnect temperature profiles on the clock skew. It is shown that new design rules must be obtained to ensure the optimal layout of the clock routing tree. This suggests that non-uniform temperature profiles along the interconnect lines can affect many steps in the design optimization flow and proper steps must be taken to ensure optimal performance. Finally, concluding remarks are made in Section 5.

2 Non-Uniform Temperature-Dependent Interconnect Delay Model

Consider an interconnect with length L and uniform width w that is driven by a driver of output resistance R_d and terminated at a load with capacitance C_L. The line is partitioned into n equal segments, each with length Δx. Using the distributed *RC* Elmore delay model, the delay D of a signal passing through the line can be written as follows:

$$D = R_d\left(\left(\sum_{i=1}^{n} c_0(x_i)\cdot\Delta x\right)+C_L\right)+\sum_{i=1}^{n} r_0(x_i)\cdot\Delta x\cdot\left(\sum_{j=i}^{n} c_0(x_j)\cdot\Delta x + C_L\right) \quad (1)$$

where $c_0(x)$ and $r_0(x)$ are the unit length capacitance and unit length resistance at location x, respectively. As the number of the partitions approaches infinity we can rewrite the Elmore delay as:

$$D = R_d\left(C_L+\int_0^L c_0(x)\,dx\right)+\int_0^L r_0(x)\cdot\left(\int_x^L c_0(\tau)\,d\tau + C_L\right)dx \quad (2)$$

The third integral in (2) represents the downstream capacitance seen by the interconnect from location x. In an interconnect experiencing

temperature profile $T(x)$ along its length, the resistance will change linearly with temperature as:

$$r_0(x) = \rho_0(1 + \beta \cdot T(x)) \quad (3)$$

where ρ_0 is the unit length resistance at reference temperature (namely 0 °C) and β is the temperature coefficient of resistance (1/°C). It is assumed that the unit length capacitance does not change with temperature variations along the interconnect length (which is generally a true assumption). It is also assumed that the temperature distribution inside the driver is uniform under the steady-state condition. Hence the R_d will be constant at the chosen operating temperature of the cell. We can simplify (2) to the following:

$$D = D_0 + (c_0 L + C_L)\rho_0 \beta \int_0^L T(x)dx - c_0 \rho_0 \beta \int_0^L x \cdot T(x)dx \quad (4)$$

where:

$$D_0 = R_d (C_L + c_0 L) + (c_0 \rho_0 \frac{L^2}{2} + \rho_0 L C_L) \quad (5)$$

D_0 is the Elmore delay of the interconnect corresponding to the unit length resistance at 0 °C. From (4) it is clear that in order to calculate the actual temperature-dependent delay we need to compute the area under $T(x)$ and $xT(x)$. To get an idea of how much temperature can affect the degradation of the delay, we assume the worst case scenario by using a uniform thermal profile at some peak temperature over the entire length of the interconnect. Choosing electrical and thermal parameters for *AlCu* interconnects with β=3E-03 (1/°C) and using r_{sh}=0.077(Ω/sq) at room temperature (25°C) and c_{sh}=0.2(fF/sq) as the unit sheet resistance and capacitance, respectively, the variations of Elmore delay with temperature in an interconnect line with w=0.32μm, R_d=10Ω, and C_L=1000fF for different lengths in μm are summarized in Figure 1.

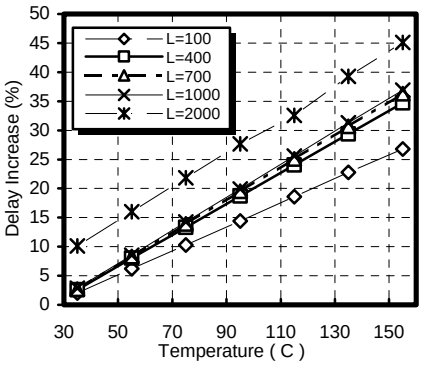

Figure 1: Percentage increase in delay with respect to the signal delay at 25 °C as a function of the line temperature.

As Figure 1 shows, for each 20-degree increase in temperature there is roughly a 5 to 6 percent increase in the Elmore delay for the long global wires. Although assuming a constant temperature along the interconnect gives an upper bound on the delay increase, we need to estimate and apply the actual variations of temperature along the interconnects in (4). This is necessary mainly due to the fact that non-uniform interconnect temperature has an unavoidable impact on the wire planning. More specifically, the non-uniform temperature profile along the interconnect can severely affect the clock skew and this effect cannot be addressed by simply accounting for a uniform worst-case maximum temperature along the interconnect.

As an example, consider having exponential temperature distributions along the interconnect length. Observing the behavior of the line under exponential thermal profiles is important in the sense that, as shown later, most of the solutions to the interconnect heat transfer equation usually have an exponential component. By applying an exponential thermal distribution $T(x)=a.exp(-bx)$ to an interconnect and using (4), the Elmore delay is as follows:

$$D = D_0 + \frac{a}{b}\rho_0 \beta[(c_0 L + C_L - \frac{c_0}{b}) + (\frac{c_0}{b} - C_L)e^{-bL}] \quad (6)$$

where D_0 is defined by (5). For the sake of analysis, consider two different exponential thermal profiles $f_1(x)$ and $f_2(x)$ along an arbitrary interconnect as depicted in Figure 2.

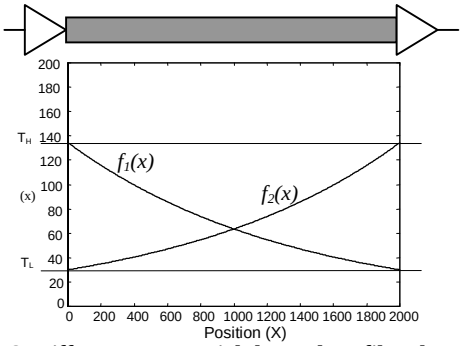

Figure 2: Different exponential thermal profiles along an interconnect line.

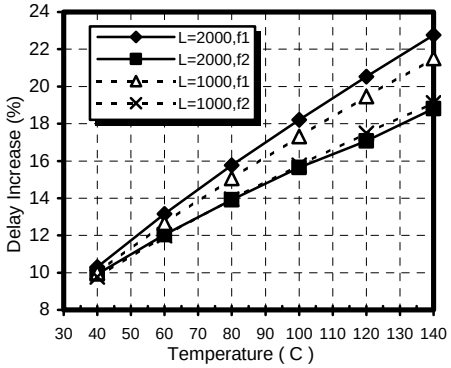

Figure 3: Performance degradation for $f_1(x)$ and $f_2(x)$ profiles of Figure 2.

Using (4), calculation shows that the interconnect Elmore delay is more adversely affected by $f_1(x)$ than by $f_2(x)$, even though the underlying areas for both $f_1(x)$ and $f_2(x)$ in Figure 2 are equal along the length of the line. Figure 3 compares the performance degradation in the presence of $f_1(x)$ and f_2(x) in two different wire lengths, 1000 μm and 2000 μm, having the same electro-thermal characteristics as mentioned before. In both cases the lower-bound temperature is kept constant at 30 °C. By increasing the upper-bound value for these functions, it can be observed that using $f_2(x)$ causes less delay increase than that caused by using $f_1(x)$. This shows that assuming a constant temperature along the wire (with peak-value) is not accurate enough in planning wire routings and clock-skew analysis, as illustrated later in more detail. The above observation demonstrates that if we have the choice, choosing thermal profile $f_2(x)$ over $f_1(x)$ is preferable. Figure 3 also demonstrates that optimizing thermal profiles is as important as minimizing interconnect length for delay optimization.

It must be noted that the substrate thermal map is strongly dependent on the design, synthesis, floorplanning and placement routines. As a result, analytical modeling of hot spots in the substrate can be a tedious task. However, to approximate a hot spot, one can assume a Gaussian thermal distribution (with constant peak temperature) along the length of a wire with median point μ at a constant peak temperature T_{max} and standard deviation σ as depicted in Figure 4. By applying $T(x) = T_{\max} \cdot \exp(-(x-\mu)^2 / 2\sigma^2)$ to (4) we can

observe the interconnect performance degradation. The movement of median μ along the length of the line will change the value of the delay degradation, and its effect on performance is also strongly dependent on the value of deviation σ. For the same σ, delay is always better for $\mu=L$ rather than for $\mu=0$ $(0 \le x \le L)$, which again shows the effectiveness of a gradual increase in the temperature along the line from source to sink. It is obvious that for the same median μ, any increase in the deviation σ will increase the delay. Figure 5 shows the increase in the delay of a wire with length 2000 μm as a function of different μ's and σ's with T_{max}=120 °C and the same electrical and thermal properties as described above for Figure 1. It can be observed that as μ moves along the line, the location at which the maximum increase in delay occurs is also a function of the deviation σ.

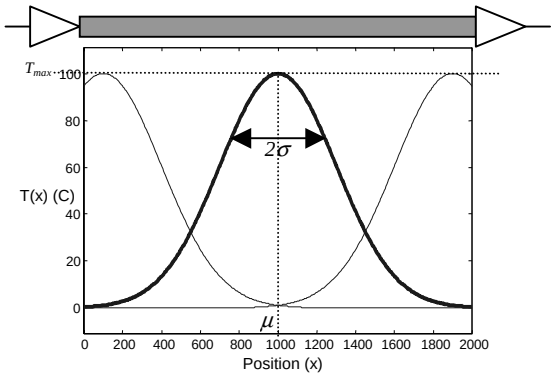

Figure 4: Constant-peak normal thermal profile with variable median μ and standard deviation σ along an interconnect line.

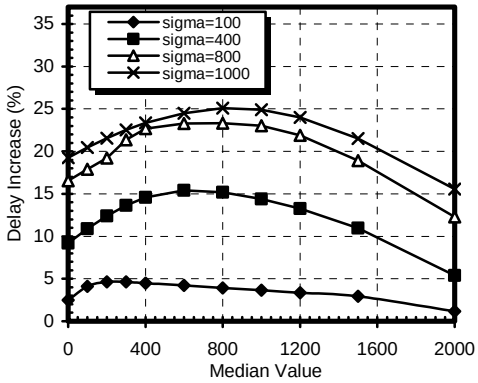

Figure 5: Delay increase as a function of the median value and the standard deviation of a normal temperature distribution.

The last two examples illustrate that the delay degradation is strongly dependent on the specific thermal distribution functions. From a resistance point of view, fluctuations of temperature along the line are equivalent to sizing a wire with uniform resistance. In sections with higher temperature, the wire is equivalent to a thinner uniform resistance wire, and in sections with lower temperature the wire acts like a thicker wire with uniform resistance. By recalling the optimization policy in uniform resistance non-uniform wire sizing [14], the best shape for such a line is a decaying exponential from the source of the signal to the destination. Considering the two previous examples of temperature profiles, when the temperature gradually increases from location *0* to *L* the line has a better performance than when there is a gradual decrease in the temperature along the length of the line. Keeping in mind that a gradual *increase* in the line temperature is equivalent to a gradual *decrease* in the size of a uniform resistance line, the results are therefore analogous to optimal uniform resistance non-uniform wire sizing (assuming a constant capacitance). Using (4), we can derive the Elmore delay for an interconnect subject to any *T(x)*. Hence the question is, what kind of *T(x)*'s usually occur along an interconnect in actual scenarios inside the chip. This issue is addressed in the next section.

3 Analytical Model for Interconnect Thermal Profile

3.1 Methodology

Heat flow in any material can be obtained by solving the fundamental heat diffusion equation in space (*x*,*y*,*z*) and time with proper boundary conditions as follows [15]:

$$\frac{\partial}{\partial x}(k\frac{\partial T}{\partial x})+\frac{\partial}{\partial y}(k\frac{\partial T}{\partial y})+\frac{\partial}{\partial z}(k\frac{\partial T}{\partial z})+Q^{*}=\delta c_{p}\frac{\partial T}{\partial t} \quad (7)$$

$$k\frac{\partial T}{\partial n_{i}}+h_{i}\cdot T=f_{i} \quad (8)$$

subject to some defined initial values. T is the time dependent temperature at each location, k is the solid thermal conductivity as a function of location and temperature (W/(m°C), Q* is the rate of heat generation per unit volume (W/m^3), δ is the solid density (kg/m^3), c_p is the specific heat (J/(kg°C)), $\partial/\partial n$ denotes the differentiation along the outward-drawn normal at the boundary surface s_i, h_i is the heat transfer function from surface s_i (W/(m^2°C)), and f_i is an arbitrary function of position.

Even though in general, the thermal conductivity k of a material is a function of temperature and position, due to its small changes in the conductors, it is usually assumed to be a constant throughout the interconnects. In addition, the four sidewalls and the top surface of the chip containing the interconnects are assumed to be insulated (which is generally a valid assumption). This means that the interconnects do not exchange energy through the four sidewalls and the top surface. The only side that can exchange heat with the interconnect is the underlying substrate (which is connected to the heat-sink). Having these assumptions and working under the steady-state condition, the system of heat equation and boundary conditions can be reduced as follows:

$$k(\frac{\partial^{2}T}{\partial x^{2}}+\frac{\partial^{2}T}{\partial y^{2}}+\frac{\partial^{2}T}{\partial z^{2}})+Q^{*}=0 \quad (9)$$

subject to the specified initial conditions. In order to find the exact solution we need to employ a 3-D finite element thermal simulation [4]. As a tradeoff, many researchers used a simplified version of (9) and employed the 1-D heat equation to avoid the huge computation time used by FEM simulators [16]. In that case (9) can be reduced as follows:

$$\frac{d^{2}T}{dx^{2}}=-\frac{Q^{*}}{k_{m}} \quad (10)$$

where k_m is the thermal conductivity of the metal. To derive the volumetric heat generation Q^*, consider an interconnect passing over the substrate as shown in Figure 6. The interconnect is connected to the substrate through vias at its two ends (not shown in Figure 6). The major source of temperature generation in a chip is the power consumption due to the switching activity of the cells lying on the substrate. In addition, the power consumption in the interconnect is also a source of the heat generation. For the interconnect line in Figure 6 the power dissipation P_g in a partial metal length Δx would be:

$$P_{g}(x)=I_{rms}^{2}\,\Delta R_{E}(x) \quad (11)$$

where I_{rms} is the root mean square part of the current passing through the line, and R_E is the electrical resistance as a function of the line temperature using (3). Initial resistance R_0 at the reference temperature can be expressed as:

$$\Delta R_0(x) = \rho_i \frac{\Delta x}{w t_m} \quad (12)$$

where ρ_i is the electrical resistivity of the interconnect at the reference temperature, t_m is the interconnect thickness and w is the width of the interconnect. On the other hand, heat energy loss due to heat transfer between the interconnect and the substrate through the insulator (oxide) in a partial length Δx is:

$$P_l(x) = \frac{T_{line}(x) - T_{ref}(x)}{\Delta R_T(x)} \quad (13)$$

where:

$$\Delta R_T(x) = \frac{t_{ox}}{k_{ox} w_{eff} \Delta x} \quad (14)$$

where $P_l(x)$ is the heat flow from the interconnect to the substrate, T_{line} is the interconnect temperature, T_{ref} is the underlying substrate temperature, R_T is the oxide thermal resistance, and k_{ox} is the oxide thermal conductivity. W_{eff} is the adjusted effective width of the interconnect to consider the heat exchange from the sides of the metal line as well as the bottom side and is approximately equal to $w(1+0.88t_{ox}/w)$ as described in [18]. Based on the above observations, the net heat energy gain per unit volume is:

$$Q^* = \frac{P_g - P_l}{w t_m \Delta x} \quad (15)$$

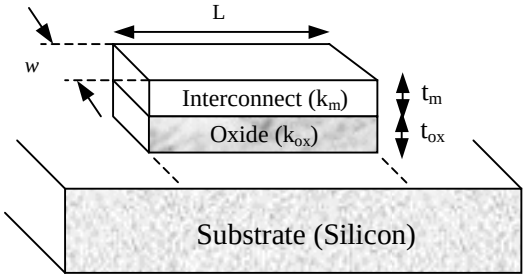

Figure 6: Interconnect line over the substrate and insulator (oxide) layer.

Using the simplified heat equation (10), the summarized interconnect heat flow equation can be written as follows:

$$\frac{d^2 T_{line}(x)}{dx^2} = \lambda^2 T_{line}(x) - \lambda^2 T_{ref}(x) - \theta \quad (16)$$

$$\lambda^2 = \frac{1}{k_m}\left(\frac{k_{ox}}{t_m t_{ox}}\left[1 + 0.88\frac{t_{ox}}{w}\right] - \frac{I_{rms}^2 \rho_i \beta}{w^2 t_m^2}\right) \quad (17)$$

$$\theta = \frac{I_{rms}^2 \rho_i}{w^2 t_m^2 k_m} \quad (18)$$

where λ and θ are constants in specified technology and interconnect layer assignment. Equation (16) and its coefficients will be the basis of our interconnect temperature calculations. Note that in order to have a unique solution for (16), we need to provide two initial conditions as well. Equation (16) shows that the underlying substrate temperature, $T_{ref}(x)$, plays an important role in determining the temperature of the line. This value is usually assumed to be constant throughout the substrate surface. Although this is a valid assumption for the short local interconnects, it is not true in the case of long global lines in the upper metal layers. Because of the different switching activities of various cells on the substrate surface, a non-uniform temperature gradient is inevitable along the substrate surface. In this analysis two cases have been studied: 1) uniform thermal profile over the underlying substrate and 2) non-uniform thermal profile over the underlying substrate.

3.2 Uniform Substrate Temperature Profile

Assume that $T_{ref}(x)$ is a constant for all positions along the length of the line. The two initial conditions that are needed to solve (16) can be derived using the interconnect and via/contact setup. For one segment of a signal net there are four possible configurations, depicted in Figure 7, based on the location and connection of the vias. Here we examine the routes between substrate and metal 1 and between metal 1 and metal 2. One can easily extend these configurations in the same manner to the other metal layers. We assume that the via will get as hot as the layer immediately beneath them. In reality (and especially in *AlCu* technology), due to their smaller cross-sectional area and higher electrical resistivity, vias can get much hotter [19] (unless they have been arranged in some sort of via array instead of just one via contact). In the current analysis, it is assumed that the router uses via arrays wherever it is possible.

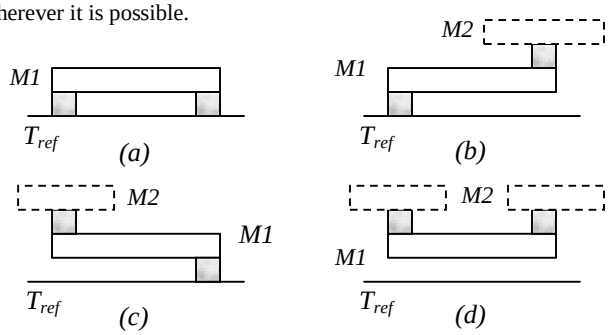

Figure 7: Different configurations of metal lines and vias.

Considering Figure 7(a), we see that the two end vias create a thermally conductive path between the metal layer and the substrate. Due to the very small thermal resistivity of vias, we assume that the temperature at the two sides of the metal line is equal to the temperature of the substrate. For example, the initial conditions in Figure 7(a) to solve (16) can be written as follows:

$$T(x=0) = T_{ref} \quad , \quad T(x=L) = T_{ref} \quad (19)$$

where $0 \le x \le L$ and T_{ref} is the constant substrate temperature. By solving the homogenous differential equation (16) with constant coefficients given by (17) and (18), the line temperature can be written as follows:

$$T(x) = T_{ref} + \frac{\theta}{\lambda^2}\left(1 - \frac{\sinh\lambda x + \sinh\lambda(L-x)}{\sinh\lambda L}\right) \quad (20)$$

Assuming a uniform substrate temperature of 100 °C, the interconnect thermal profile for global lines corresponding to Figure 7(a) for two different technologies (where parameters are taken from ITRS [17]) are depicted in Figure 8. Distance d is defined as the heat diffusion length; it is a function of 1/λ and is strongly dependent on the thickness of the oxide between the metal and the substrate and the effective current density flowing through the metal. Using (17) and assuming a constant current density in all metal layers of a signal net, the diffusion length d is larger for the higher level metal layers due to their higher underlying oxide thickness. As an example, for an interconnect with an RMS current of 2*mA* in a metal layer with width 0.32 μm and an underlying oxide layer with thickness 1.2 μm, the diffusion length d is approximately 40 μm. In addition, the peak value of the temperature in Figure 8 is equal to θ/λ^2. For interconnects whose lengths are comparable to the heat diffusion lengths, the line temperature does not reach the maximum peak value. Using this concept, the authors in [20] have introduced a new technique to make the peak temperature lower by adding extra dummy vias separated by a distance less than the diffusion length.

By investigating the effect of the temperature profile in Figure 8 on the interconnect delay equation given in (4), it is clear that the delay has a strong dependency on the value of λ. Although decreasing λ

increases the value of the diffusion length, for a long global line, it also increases the peak value (θ/λ^2) sharply and, consequently, increases the Elmore delay. It can also be shown that the delay of interconnect configuration depicted in Figure 7(b) is lower than that for Figure 7(c), due to the fact that the configuration in Figure 7(b) has a rising thermal profile while the one in Figure 7(c) has a decaying thermal profile.

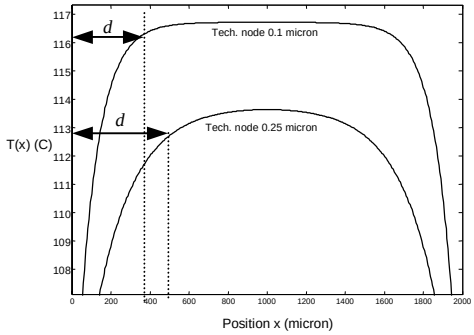

Figure 8: Thermal profile along the length of a 2000 μm long interconnect (Cu) line with uniform substrate temperature using interconnect parameters for global metal layers in the 0.1 μm and 0.25 μm technology nodes provided by ITRS.

3.3 Non-uniform Substrate Temperature Profile

Considering the different switching activities and power consumptions of the cells on the substrate, $T_{ref}(x)$ cannot be a constant value in all locations, x. The quality of extracting the substrate thermal profile depends on how accurately one estimates the power consumptions of the cells. Some techniques that are used to find the substrate thermal profile are given in [12]. Due to the duality between thermal and electrical networks, the easiest way to map the substrate thermal profile is to model the substrate as a 3-D grid and solve the system of thermal equations between every two nodes while considering the packaging and the ambient temperature as additional thermal nodes. However, because this procedure depends on finding the power map of the substrate, $T_{ref}(x)$ is a design dependent function. For this reason, and for illustration, we use two linear substrate temperature distributions along the length of an interconnect and observe their effects on interconnect $T(x)$ variations. First, we use $T_1(x)=ax+b$ and solve the non-homogeneous differential heat equation (16) for the configuration shown in Figure 7(a) with proper initial conditions. The resulting thermal profile along the line can be expressed as:

$$T_1(x) = \frac{\theta}{\lambda^2}(1-e^{-\lambda x} - \frac{1-e^{-\lambda L}}{\sinh \lambda L}\sinh \lambda x) + ax + b \qquad (21)$$

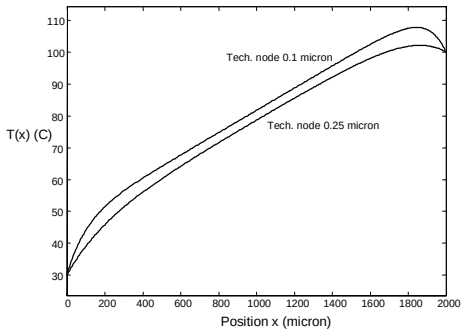

Figure 9: Thermal profile along the length of a 2000 μm long interconnect (Cu) line with a linear substrate thermal profile using parameters of global wires of 0.1 μm and 0.25 μm technologies provided by the ITRS.

In the same way, by using another substrate temperature profile along the length of the interconnects, $T_2(x)=a(L-x)+b$, which is the same linear profile in the opposite direction along the length of the line, one can calculate the line temperature profile. Figure 9 shows the thermal profile in an interconnect using the linear substrate thermal profile $T_1(x)$ (with a gradient from 30 °C to 100 °C). After calculating the Elmore delay in presence of these two thermal profiles by using (4), it can be shown that the delay degradation due to $T_2(x)$ is worse than that due to $T_1(x)$. This result is the same as the computed results in Section 2 and illustrates the importance of the directional gradients in the substrate temperature profile.

4 Impact of Non-uniform Interconnect Temperature on Clock Skew

As shown in Section 2, the increase in the Elmore delay can be significant at high temperatures. Moreover, delay variations arising from non-uniform interconnect thermal profiles cannot be accounted for by estimating a worst-case delay based on a uniform maximum temperature along the wires. Consequently, a serious problem may arise, which is the skew fluctuations in a clock signal net. This may in turn degrade the performance of the circuit. Assume a clock net with two fanouts as illustrated in Figure 10. For simplicity assume that both wires 1 and 2 have the same lengths, widths, and electro-thermal characteristics (used in Section 2) and are routed on the same layer.

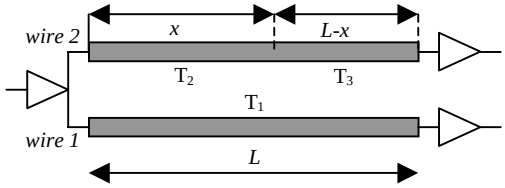

Figure 10: Portion of a clock tree with two fanout branches that have equal wire length.

Assuming different but uniform temperature profiles along both wires, the signal skew can be extracted from Figure 1 by estimating the difference in delay corresponding to the two uniform temperature profiles. A more realistic case arises if one of the wires develops a non-uniform thermal profile along its length due to some underlying thermal gradients over the substrate. In the worst case, we can assume that a section of the line is at one temperature and the rest of the line is at another temperature, as shown in Figure 10 (for wire 2), with the length x at temperature T_2 and the length $(L-x)$ at temperature T_3. Figure 11 depicts the percentage of the normalized delay increase between wires 1 and 2 as a function of position x in which the thermal gradient occurs at location x, wire 1 is at a uniform temperature of 100 °C, and both the wires are 2000 μm in length. It can be observed that as x approaches zero, the percentage of delay increase reaches its maximum value since the hotter section of the wire (L-x) (which is at T_3) extends over the entire length of the line.

Now assume that with wire 1 remaining at temperature T_1, wire 2 has a certain section of fixed length x where the temperature is lower (or higher) than the rest of the wire. We proceed to study the effect of the magnitude of the gradient between these two sections x and L-x in wire 2 on the normalized delay difference. Assume that temperature T_2 in section x of wire 2 is at uniform temperature of 80 °C while wire 1 is still at uniform temperature of 100 °C. Figure 12 shows that the percentage of normalized delay difference between wires 1 and 2 is a function of the magnitude of the temperature gradient in wire 2. It can also be observed that the magnitude of the thermal gradient is an important factor in the signal skew fluctuations. In this example, due to the specific definition of the thermal gradient, skew becomes zero in a certain location along the length of the wire.

The above analysis shows the importance of considering the effects of the non-uniform interconnect temperature on the clock skew. Due to the high currents driven through the clock wires, clock nets usually exhibit the highest Joule heating among signal nets, and since they span a large area over the die, the probability that they will experience some thermal gradients is much higher than that for the shorter signal nets. As a result, careful consideration of non-uniform temperature profiles is necessary in clock skew estimation along the clock signal net. It can be shown that by using (4) and having the non-uniform thermal profiles along wires 1 and 2, one can calculate the effective ratio of the length of wire 1 to that of wire 2 such that the signal skew is eliminated. This design rule may be used in the bottom-up merging clock tree generation techniques (such as H-tree technique) to ensure a near-zero clock skew routing in the presence of substrate thermal non-uniformities [21].

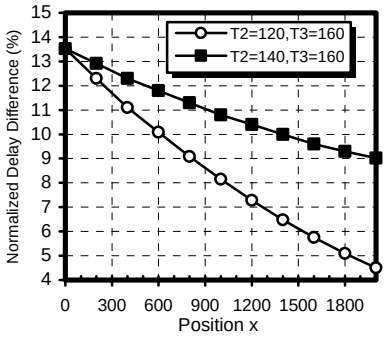

Figure 11: Percentage of normalized delay difference between wires 1 and 2 as a function of location parameter *x*.

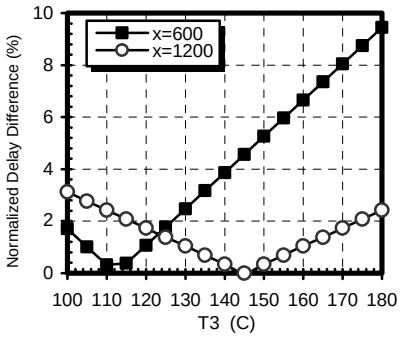

Figure 12: Percentage of normalized delay difference between wires 1 and 2 as a function of the temperature T_3 as shown in Figure 10.

5 Conclusions

It was shown that non-uniform temperature distributions along long global wires in high-performance ICs could have significant implications for interconnect performance and other critical design metrics such as the clock skew. A detailed analysis of the impact of non-uniform temperature distributions on the interconnect performance was presented using a new distributed *RC* delay model that incorporates non-uniform interconnect temperature dependency. The model was applied to analyze a wide variety of interconnect layouts and temperature profiles. Analytical models for accurate interconnect temperature distributions arising from non-uniform substrate temperature profiles were derived using fundamental heat diffusion equations. It was shown that the clock skew would be significantly affected by interconnect temperature non-uniformities. These studies suggest that incorporation of thermal analysis is necessary in performing various design optimization steps in high performance ICs.

Acknowledgements

This work was supported in part by the SRC under contract number 98-DJ-606. The authors would also like to acknowledge support from the MARCO Interconnect Focus Center at Stanford University.

6 References

[1] L.Gwennap, "Power issues may limit future CPUs," *Microprocessor Report*, 10(10), August 1996.

[2] Y-K.Cheng *et al*, "ILLIADS-T: an electrothermal timing simulator for temperature-sensitive reliability diagnosis of CMOS VLSI chips," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 8, pp.668-681, 1998.

[3] K.Banerjee *et al*, "On thermal effects in deep sub-micron VLSI interconnects," *Proc. Design Automation Conference*, 1999, pp. 885-891.

[4] S.Rzepka *et al*, "Characterization of self-heating in advanced VLSI interconnect lines based on thermal finite element simulation," *IEEE Transactions on Components, Packaging and Manufacturing Technology-A*, vol. 21, No. 3, pp. 406-411, 1998.

[5] J.R.Black, "Electromigration- A brief survey and some recent results," *IEEE Transactions on Electron Devices*, vol. ED-16, pp.338-347, 1969.

[6] Y.K.Cheng *et al*, "iCET: A Complete Chip-Level Thermal Reliability Diagnosis Tool for CMOS VLSI Chips," *Proc. Design Automation Conference*, 1996, pp.548-551.

[7] J.Tao *et al*, "Electromigration Design Rules for Bidirectional Current", *Proc. Int'l Reliability Physics Symp.*, 1996, pp.180-187.

[8] S.Im, K.Banerjee, "Full chip thermal analysis of planar (2-D) and vertically integrated (3-D) high performance ICs," *Tech. Digest IEDM*, 2000, pp.727-730.

[9] Y.K.Cheng *et al*, *Electrothermal Analysis of VLSI Systems*, 1st ed., Kluwer Academic Publishers, 2000.

[10] Z.Yu, D.Yergeau, R.W. Dutton, S.Nakagawa, N. Chang, S. Lin and W.Xie, "Full Chip Thermal Simulation," *Proceedings of the 1st IEEE International Symposium on Quality Electronic Design*, 2000, pp.145-149.

[11] Q.Wu, Q.Qiu, M.Pedram, "Dynamic Power Management of Complex Systems using Generalized Stochastic Petri Nets," *Proc. Design Automation Conference*, 2000, pp.352-356.

[12] C.H.Tsai, S.M.Kang, "Cell-Level Placement for Improving Substrate Thermal Distribution," *IEEE Trans. on Computer Aided Design*, Vol 19, No 2, pp 253-265, 2000.

[13] C.Tsai, S.M.Kang, "Fast Temperature Calculation for Transient Electrothermal Simulation by Mixed Frequency/Time Domain Thermal Model Reduction," *Proc. Design Automation Conference*, 2000, pp. 750-755.

[14] C-P.Chen, Y-P.Chen, D.F.Wong, "Optimal Wire-Sizing Formula Under the Elmore Delay Model," *Proc. Design Automated Conference*, 1996, pp.487-490.

[15] A.J.Chapman, *Fundamentals of Heat Transfer*, 4th ed., New York, Mcmillan, 1984.

[16] H.A.Schafft, "Thermal Analysis of Electromigration Test Structures," *IEEE Trans. on Electron Device*, vol. Ed-34, No.3, pp.664-672, 1987.

[17] The International Technology Roadmap for Semiconductors (ITRS), 1999 Edition.

[18] A.A.Bilotti, "Static Temperature Distribution in IC Chips with Isothermal Heat Sources," *IEEE Transaction on Electron Device*, Vol. ED-21, No. 3,pp.217-226, 1974.

[19] M.Igeta *et al*, "Thermal Characteristics of Submicron Vias Studied by Scanning Joule Expansion Microscopy," *IEEE Electron Device Letters*, vol.21, No.5, pp.224-226, 2000.

[20] T.Y.Chiang, K.Banerjee, and K.C.Saraswat, "Effect of Via Separation and Low-k Dielectric Material on the Thermal Characteristics of Cu Interconnects," *Technical Dig. IEDM*, 2000, pp.261-264.

[21] A.H.Ajami, M.Pedram, K.Banerjee, "Effects of non-uniform substrate temperature on the clock signal integrity in high performance designs," *Proc. IEEE Custom Integrated Circuits Conference*, 2001.