Modeling and Analysis of Non-Uniform Substrate Temperature Effects on Global ULSI Interconnects

Amir H. Ajami[1*], *Member,* Kaustav Banerjee[2], *Senior Member,* and Massoud Pedram[3], *Fellow*

[1]Magma Design Automation, Santa Clara, CA, 95054 (amira@magma-da.com)

[2]Dept. of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 (kaustav@ece.ucsb.edu)

[3]Dept. of EE-Systems, University of Southern California, Los Angeles, CA 90089 (pedram@ceng.usc.edu)

Abstract

Non-uniform thermal profiles on the substrate in high-performance ICs can significantly impact the performance of global on-chip interconnects. This paper presents a detailed modeling and analysis of the interconnect performance degradation due to the non-uniform temperature profiles that are encountered along long metal interconnects as a result of existing thermal gradients in the underlying Silicon substrate. A non-uniform temperature-dependent distributed *RC* interconnect delay model is proposed. The model is applied to a wide variety of interconnect layouts and substrate temperature distributions to quantify the impact of such thermal non-uniformities on signal integrity issues including speed degradation in global interconnect lines and skew fluctuations in clock signal distribution networks. Subsequently, a new thermally dependent zero-skew clock routing methodology is presented. This study suggests that thermally-aware analysis should become an integrated part of the various optimization steps in physical-synthesis flow to improve the performance and integrity of signals in global ULSI interconnects.

[*] This work has been done while the author was with Dept. of EE-Systems, University of Southern California.

Keywords: clock skew, Elmore Delay, global interconnects, hot-spots, on-chip temperature variations, signal integrity, substrate thermal gradient.

1 Introduction

CMOS device technology has scaled rapidly for nearly three decades and has come to dominate the electronics world. Because of this scaling, CMOS circuits have become extremely dense and operate at ever-increasing clock frequencies. Consequently, power dissipation and thermal issues have become major design considerations for high-performance ICs including microprocessors. Although industry projections predict for at least another 10 years of progress, making progress will be difficult, and will likely be significantly constrained by power dissipation and heat generation inside the chips. Thermal issues are rapidly becoming one of the most challenging problems in high-performance IC design due to aggressive device scaling trends [1],[2],[3]. Hence, thermal management has become critical in the design and development of future generations of high-performance microprocessors, integrated network processors, and systems-on-a-chip.

At the circuit level, thermal effects have important implications for both performance and reliability of the chip [4],[5],[6]. In general, on-chip temperature rise is known to increase transistor and interconnect delays and thereby degrade circuit performance. The chip temperature rise is also known to degrade IC reliability since most reliability mechanisms have strong temperature dependence. In particular, interconnect performance and reliability factors are known to degrade with increase in metal temperature. The interconnect reliability degradation is mainly due to the electromigration (EM) phenomenon [7], although Joule heating (self-heating) in global lines is an additional contributor to the interconnect reliability degradation [8]. The interconnect performance degradation is primarily due to the fact that the resistivity of metal interconnect increases linearly with its temperature, thereby increasing interconnect resistance (and consequently the signal propagation delay). In addition, the increase in the resistivity of the interconnect leads to an increase in interconnect Joule heating that causes an additional temperature rise in the interconnect [5].

Although extensive work has been done to determine the chip temperature and predict its impact on EM reliability of interconnects [4],[5],[6],[8],[9],[10], few efforts have focused on analyzing the effect of temperature on the performance of interconnects. Recent work indicates that, in high performance ICs, the peak chip temperature can raise up to 160 ºC in 90nm technology node and is expected to rise even

further for future technology nodes [11]. Since, these peak temperatures always occur at the top of the chip, they can significantly increase the interconnect resistance, which would in turn increase the signal propagation delay in the interconnect lines. In some cases, this effect has also been shown to cause timing violation [12]. However, all of the previous works on timing analysis have assumed a *uniform* temperature profile of the substrate. This assumption is in general invalid and becoming increasingly inaccurate for nanometer scale high-performance ICs, where large temperature gradients can occur in the substrate. These gradients, for example, may be created due to "spotty" gate-level switching activity and/or because various functional blocks are put in different operational modes, e.g., active, standby, or sleep modes [13]. Dynamic power management (DPM) [14], clock gating, and non-uniform gate-level switching activity among different logical blocks can be major contributors to a non-uniform substrate temperature. In fact, it has been recently reported that thermal gradients as large as 50 °C can exist across high-performance microprocessor substrates [15],[16]. Based on the cell-level power consumption map of the substrate, researchers have proposed efficient techniques to obtain temperature profile of the substrate surface [17],[18].

It is widely known that as the minimum feature size of the CMOS fabrication process continues to scale down, the performance of the CMOS integrated circuits has become dominated by the global interconnect lines [19]. The existence of thermal gradients on the substrate results in non-uniform temperature profiles along the length of these global interconnect lines running above the substrate, which in turn leads to non-uniform resistance profiles for these interconnect lines. Such non-uniformity in the resistance of global interconnects strongly impacts many aspects of interconnect performance modeling and optimization. In addition, as feature size scales to sub-100nm dimensions, in spite of an increase in the number of metal layers that will be available in advanced technology nodes, the top metal layers may get closer to the substrate, which will result in stronger thermal coupling between the substrate and the interconnect lines, and thereby impact interconnect performance analysis. Clearly, the dependence of interconnect performance on non-uniform temperature distributions along the length of global wires will have a big impact on the solutions to many physical design and layout optimization problems, including clock skew control, wire sizing, layer assignment, crosstalk effects, and buffer insertion. These observations suggests that non-uniform temperature profiles along the interconnect lines should be considered during the design optimization flow and proper steps should be taken to ensure the optimal performance [20]. It should be noted that the effects of such thermal gradients on the performances of devices are also very important and perhaps more severe than such effects on the interconnect performance. However, it is important to separately analyze the implications of these thermal gradients on interconnects and devices. Hence, this paper focuses on the impact of substrate thermal gradients on the

interconnect performance. Implications of non-uniform substrate temperature for device performance can be found in [21].

This paper presents a systematic study of the impact of non-uniform substrate thermal profiles and their impact on interconnect delay and signal integrity in general, and clock skew in particular. The paper is organized as follows: Section 2 provides the necessary background for heat conduction analysis and provides a quantitative way of estimating the temperature along the length of an interconnect line in the presence of substrate thermal gradients. Meaningful boundary conditions and interconnect-via/contact arrangements encountered in current chip designs are used to obtain the thermal profiles along the interconnect lines. Section 3 introduces the modeling methodology of the impact of non-uniform interconnect thermal profile on the Elmore delay. A distributed *RC* interconnect delay model as a function of the non-uniform line temperature is proposed, and a design methodology for improving the performance in the presence of non-uniform interconnect thermal profiles is presented. Section 4 illustrates the effects of non-uniform interconnect temperature profiles on the clock skew. New design rules are proposed to ensure near-optimal layout needed for zero-skew clock tree. Finally, concluding remarks are made in Section 5.

2 Analytical Model for Estimating Interconnect Thermal Profile

2.1 General Theory

The temperature distribution as a function of position (*r*) and time (*t*) in a solid is governed by the following heat diffusion equation:

$$-\nabla \cdot (-k\nabla T(t,r)) + Q(r) = \frac{d}{dt} c_p T(t,r) \qquad (2.1)$$

subject to specific initial values and proper boundary conditions. T is the time- and position-dependent temperature at each co-ordinate r, k is the solid thermal conductivity of the material as a function of temperature (W/(m°C)), c_p is the specific heat (J/(kg°C)) of the material constituting the structure, and Q is the heat generation rate. In a general multi-layer structure, k and Q are position-dependent, i.e., they are functions of r. In a 3-D space (x,y,z), the heat diffusion equation (2.1) in any material can be written as [22]:

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial T}{\partial z}\right) + Q^{*} = \delta\, c_p \frac{\partial T}{\partial t} \qquad (2.2)$$

where Q^* is the rate of heat generation per unit volume (W/m^3), δ is the solid density (kg/m^3). A general boundary condition for solving the diffusion equation (2.2) can be written as follows:

$$k\frac{\partial T}{\partial n_i} + h_i \cdot T = f_i \tag{2.3}$$

$\partial/\partial n$ 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 in the space.

Although the thermal conductivity k is generally a function of temperature and position, due to its rather small variations in the conductors, k is often assumed to be a constant during analysis of ULSI interconnects. In addition, the four sidewalls and the top surface of the chip containing the interconnect lines are presumed to be completely insulated (which is generally a valid assumption [11]). This means that the interconnect lines can exchange heat with the environment only through the bottom face, i.e., the chip substrate, which is in turn, connected to the heat-sink through the packaging material. Based on these simplifying assumptions and working under the steady-state condition, the system of heat equation (2.2) and boundary conditions can be reduced as follows:

$$k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) + Q^*_{eff} = 0 \tag{2.4}$$

subject to a set of specified initial conditions. Note that Q^*_{eff} is the *effective* volumetric heat generation, which also considers the heat *loss* rate per unit volume that addresses the functionality of the boundary condition (heat loss) from the bottom side and side surfaces of the interconnect line. In order to find an exact solution of equation (2.4) we need to employ a 3-D finite element thermal simulation [6]. On the other hand, for a global interconnect line, the length of the line is sufficiently larger than its thickness or width, i.e., the thermal gradients along the thickness and width of the interconnect line can be ignored when studying those long interconnects. Consequently, many researchers have used a simplified version of (2.4) and employed the 1-D heat diffusion equation to avoid the extensive computation time needed by FEM simulators while marinating acceptable results [8]. In that case, equation (2.4) can be reduced as follows:

$$\frac{d^2 T}{dx^2} = -\frac{Q^*_{eff}}{k_m} \tag{2.5}$$

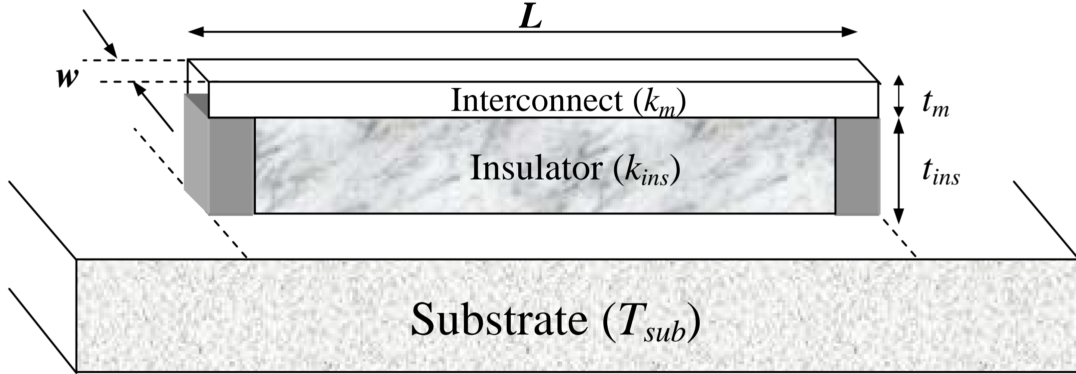

Figure 1. An interconnect line passing over the substrate, separated by an insulation layer.

where k_m is the thermal conductivity of the metal. To derive the effective volumetric heat generation Q^*_{eff}, consider an interconnect line passing over the substrate surface as shown in Figure 1. The interconnect line is connected to the substrate through vias at its two ends. The major source of temperature generation in a chip is the power dissipation due to the dynamic and static activity of the cells lying on the substrate. In addition, the power dissipation in the interconnect line is also a source of heat generation.

For the interconnect line shown in Figure 1, the power dissipation P_g in a partial metal length Δx can be expressed as:

$$P_g(x) = I_{rms}^2 \Delta R_E(x) \tag{2.6}$$

where I_{rms} is the root mean square current passing through the line. The electrical resistance of the interconnect line R_E has a linear relationship with its temperature and can be written as follows:

$$R_E(x) = R_0(1 + \beta \cdot T(x)) \tag{2.7}$$

where R_0 is the resistance per unit length at a reference temperature, β is the temperature coefficient of resistance (1/°C), and $T(x)$ is the temperature profile along the length of the interconnect line. Furthermore, initial resistance R_0 can be expressed as:

$$\Delta R_0(x) = \rho_i \frac{\Delta x}{w t_m} \qquad (2.8)$$

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, energy loss due to the heat transfer between the interconnect and substrate through the insulator for a partial length Δx is:

$$P_l(x) = \frac{T_{line}(x) - T_{sub}(x)}{\Delta R_T(x)} \qquad (2.9)$$

where:

$$\Delta R_T(x) = \frac{t_{ins}}{k^*_{ins} w_m \Delta x} \qquad (2.10)$$

$P_l(x)$ is the heat flow from the interconnect to the substrate, T_{line} is the interconnect temperature, T_{sub} is the underlying substrate temperature, R_T is the insulator thermal resistance, and k^*_{ins} is the *effective* insulator thermal conductivity. k^*_{ins} is a shape-dependent parameter which considers the geometrical configuration of the heat conducting body on the thermal conductivity. In the case of heat flow by conduction between two identical flat plates with insulated edges, k^*_{ins} is simply the thermal conductivity k_{ins}. For the case of a rectangular shape parallel to an infinite plate, a simple approximation introduced by Bilotti [23] can be used. This approximation uses a quasi 2-D model where heat flow path assumed to be through the bottom side and partially through the two sides along the length of a rectangular shape (i.e. interconnect line). In this case, the effective thermal conductivity k^*_{ins} can be expressed as $k_{ins}((1+0.88t_{ins}/w)$ and provides results with 3% accuracy for test cases with $w_m/t_{ins} > 0.4$. However, in deep submicron technologies, the geometrical dimensions of global lines do not satisfy this condition. Hence, using Bilotti's estimation for effective thermal conduction often results in somewhat higher values for the peak temperature in global lines compared to the actual values. In reality, the heat flows from all sides of the rectangular body (i.e., interconnect line). A more accurate expression for k^*_{ins} was given in [24] where it takes this factor into account and therefore provides a more accurate expression for the thermal conductivity as follows:

$$k^*_{ins} = k_{ins} \cdot \frac{t_{ins}}{w_m} \cdot 1.685 \cdot [\log(1 + \frac{t_{ins}}{w_m})]^{-0.59} \cdot (\frac{t_{ins}}{t_m})^{-0.078} \qquad (2.11)$$

Authors in [25] have used this approximation and validated its accuracy with 3-D FEM simulations with negligible error for rectangular-shaped interconnect lines. Based on the above observations, the net heat energy gain per unit volume is:

$$Q^{*}_{eff} = \frac{P_g - P_l}{w t_m \Delta x} \tag{2.12}$$

Using the simplified heat equation (2.5), the summarized interconnect heat flow equation can be written as follows:

$$\frac{d^2 T_{line}(x)}{dx^2} = \lambda^2 T_{line}(x) - \lambda^2 T_{sub}(x) - \theta \tag{2.13}$$

$$\lambda^2 = \frac{1}{k_m}\left(\frac{k^{*}_{ins}}{t_m t_{ins}} - \frac{I^2_{rms}\rho_i \beta}{w^2 t_m^2}\right) \tag{2.14}$$

$$\theta = \frac{I^2_{rms}\rho_i}{w^2 t_m^2 k_m} \tag{2.15}$$

where λ and θ are constants in a specific technology node and interconnect layer assignment. Equation (2.13) and its coefficients will be the basis of our interconnect temperature calculations. Note that in order to have a unique solution for (2.13), two initial conditions should be provided. Equation (2.13) shows that the underlying substrate temperature, $T_{sub}(x)$, plays an important role in determining the temperature of the line. This profile 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. As mentioned earlier, because of the different switching activities of various blocks on the substrate surface, a non-uniform temperature profile along the substrate surface is inevitable. In this study two cases have been analyzed: 1) uniform thermal profile over the underlying substrate and 2) non-uniform thermal profile over the underlying substrate.

2.2 Effect of Uniform Substrate Temperature on Interconnect

Assume that $T_{sub}(x)$ is constant for all positions along the length of the line. The two initial conditions that are needed to solve (2.13) can be derived using the interconnect line and the via/contact setup. For one segment of a signal net there are four possible configurations, depicted in Figure 2, based on the location and connection of the vias. Here the routes between substrate and metal 1 and between metal 1 and metal 2 are examined. One can easily extend these configurations in the same manner to the other

metal layers. It is assumed that vias get as hot as the layers that are immediately beneath them. In reality (and especially in *AlCu* technology that employs Tungsten vias), due to their smaller cross-sectional area and higher electrical resistivity, via's can become much hotter [26] (unless they have been arranged in an array format instead of just one via). In the present analysis, it is assumed that the router uses via arrays wherever it is possible.

Considering Figure 2(a), it can be seen 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, it is assumed that the temperature at the two sides of the metal line is equal to the temperature of the substrate. The initial conditions in Figure 2(a) to solve (2.13) can be written as follows:

$$T(x=0)=T_{sub} \quad , \quad T(x=L)=T_{sub} \tag{2.16}$$

where $0 \le x \le L$ and T_{sub} is the constant substrate temperature. By solving the homogenous differential equation (2.13) with constant coefficients given by (2.14) and (2.15), the interconnect line thermal profile can be written as follows:

$$T(x)=T_{sub}+\frac{\theta}{\lambda^2}\left(1-\frac{\sinh\lambda x+\sinh\lambda(L-x)}{\sinh\lambda L}\right) \tag{2.17}$$

Assuming a uniform substrate temperature of 100 °C, the interconnect thermal profile for the global lines corresponding to Figure 2(a) for two different technologies (with parameters provided by NTRS [27]) are depicted in Figure 3. Distance *d* is defined as the heat diffusion length; it is a function of $1/\lambda$ and is strongly dependent on the thickness of the insulator between the metal and the substrate and the effective current density in the metal line. Using (2.17) 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 insulator 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 line temperature in Figure 3 is equal to θ/λ^2. For interconnects whose lengths are comparable to the heat diffusion length, the line temperature does not reach the estimated maximum peak value. Using this concept, there are new techniques to make the peak temperature lower by adding extra dummy vias separated by a distance less than the thermal diffusion length [28].

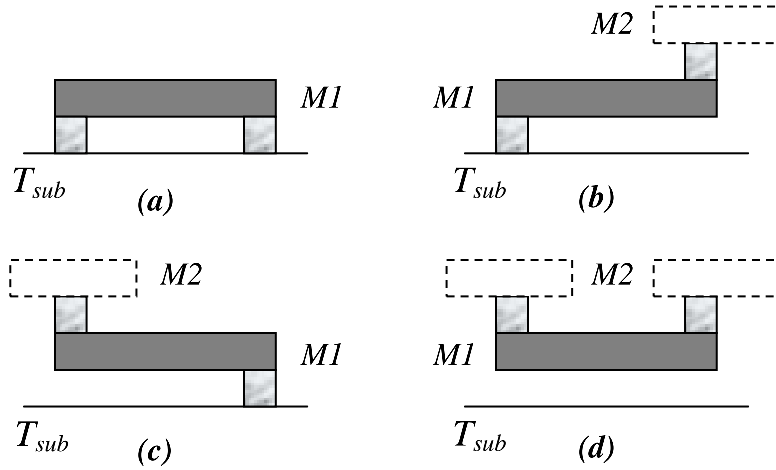

Figure 2. Different configurations of interconnect metal lines at layers M_1 and M_2 and the vias connected at their two ends, in presence of an underlying substrate at temperature T_{sub}.

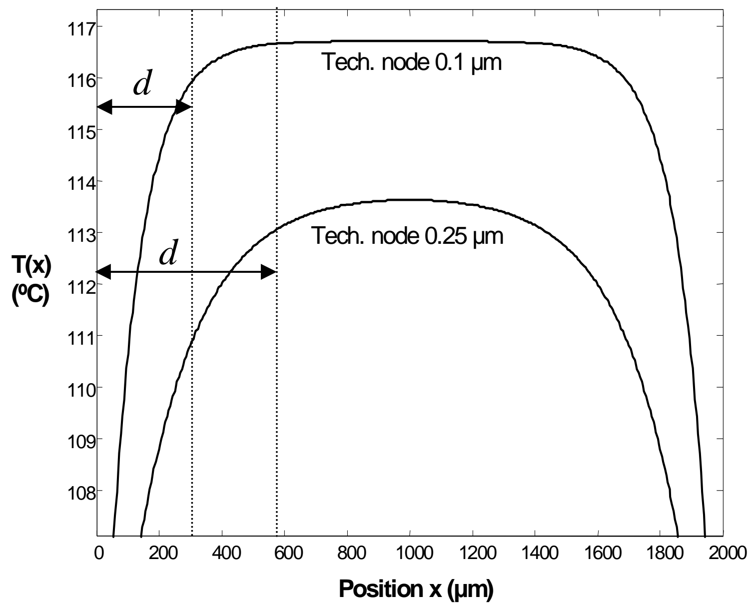

Figure 3. Thermal profile along the length of a 2000 µm long interconnect line (Cu) carrying a constant current with a uniform substrate temperature subject to electro/thermal interconnect parameters for top-most global metal layer in 0.1 µm and 0.25 µm technology nodes provided by NTRS [27]. Distance *d* is the heat diffusion length.

Using parameters provided by ITRS [29] for sub-100nm technology, it has been predicted that in the worst-case scenario, the maximum interconnect temperature in a global line may reach up to 210 °C for 50nm technology node [11].

2.3 Effect of Non-uniform Substrate Temperature Profile on Interconnect

Considering different switching activities and power consumptions of the cells, substrate thermal profile $T_{sub}(x)$ may not be a constant function in all locations, x, on the substrate surface. The quality of extracting the $T_{sub}(x)$ depends on how accurately one estimates the power consumptions of the cells or macro-cells. Some techniques that are used to find the substrate thermal profiles are outlined in [17]. 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 resistive grid and solve the system of thermal relations between every two adjacent nodes in the grid while considering the packaging and the ambient temperature as additional thermal nodes (a more simplistic technique is to use a 2-D mesh over the substrate surface, c.f. Figure 4). This can be realized by using the concept of *transfer thermal resistance*. By definition, the transfer thermal resistance $R^T{}_{ij}$ of a location j (in the 2-D mesh or 3-D grid) with respect to a point heat source i can be defined as:

$$R^T{}_{ij} = \frac{T_{ji}}{P_i} \tag{2.18}$$

which is basically the dual of the electrical relationship between current in an electrical resistor and the voltages at its two ends. Using the finite difference method [17], one can easily find the transfer thermal resistance values of all surface nodes with respect to any single source node by sampling the temperature of these nodes due to one unit of dissipated heat at the source node. In general, by using a 3-D grid structure over the substrate and (2.18), one could formulate an $n \times n$ thermal resistance matrix for 3-D nodes, where n is the total number of nodes in the grid. Using the transfer thermal resistance matrix $\mathbf{R}^T{}_{n\times n}$, one can easily calculate the temperature distribution $\mathbf{T} = [T_1, T_2, \ldots, T_n]^t$ at each node of the grid due to a given specific heat distribution $\mathbf{P} = [P_1, P_2, \ldots, P_n]^t$, by solving $\mathbf{T} = \mathbf{R}^t \times \mathbf{P}$.

As this procedure depends on finding the power map of the cells on the substrate, $T_{sub}(x)$ is a design dependent function. For this reason, and for illustration, we use a linear substrate temperature distribution along the length of an interconnect line and observe its effects on interconnect temperature $T(x)$ variations.

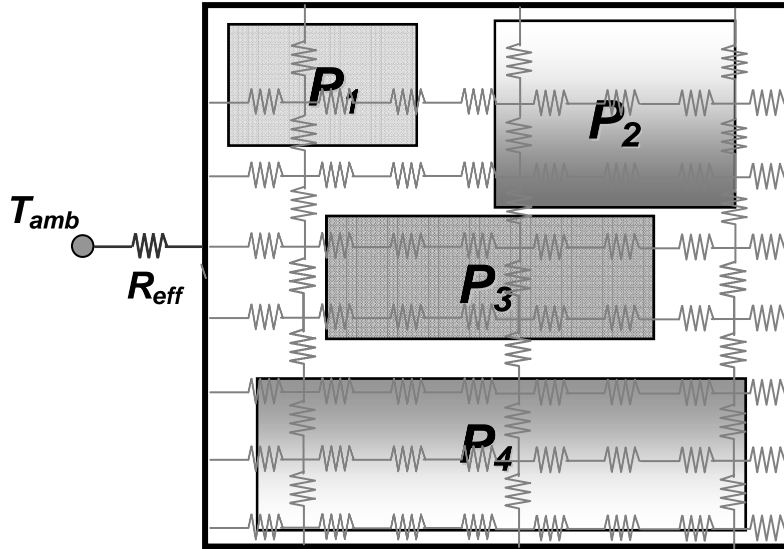

Figure 4. Illustrating the concept of imposing a 2-D thermal resistive mesh on the substrate surface for determining $T_{sub}(x)$, and using the thermal resistance between adjacent nodes in the mesh by considering the non-uniform power consumption of each block. One can extend this technique to a 3-D grid model and solve the system of linear equations to derive the temperature of each node in the grid. The constant ambient temperature T_{amb} has been simply modeled by an extra thermal node connected to all the boundary nodes in the mesh through an effective thermal resistance R_{eff} which models the packaging and heat sink thermal path.

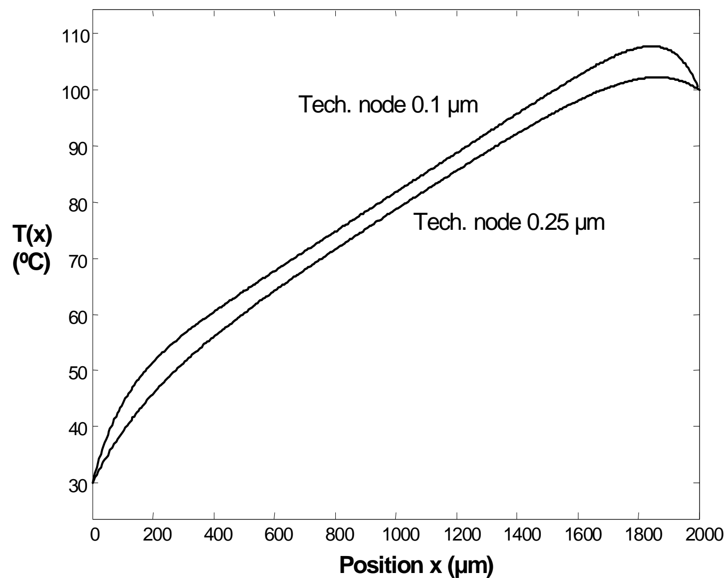

Figure 5. Thermal profile along the length of a 2000 µm long interconnect line (Cu) carrying a constant current with a linear substrate thermal profile (T_L=30, T_H=100), subject to electro/thermal interconnect parameters of top-most global metal layer at 0.1µm and 0.25µm technology nodes provided by the NTRS [27].

Lets assume $T_1(x) = ax+b$ and solve the non-homogeneous differential heat equation (2.14) for the configuration shown in Figure 2(a) with proper initial conditions. The resulting thermal profile along the line can be expressed as:

$$T_1(x) = \frac{\theta}{\lambda^2}(1 - e^{-\lambda x} - \frac{1 - e^{-\lambda L}}{\sinh \lambda L} \sinh \lambda x) + ax + b \qquad (2.19)$$

Figure 5 shows the thermal profile in an interconnect using the linear substrate thermal profile $T_1(x)$ (by applying an artificial thermal gradient from 30 °C at one side of the line to 100 °C at the other side).

3 A Non-Uniform Temperature-Dependent Interconnect Delay Model

Consider an interconnect with length L and uniform width w that is driven by a driver with on-resistance R_d and junction capacitance C_p terminated by a load with capacitance C_L as depicted in Figure 6.

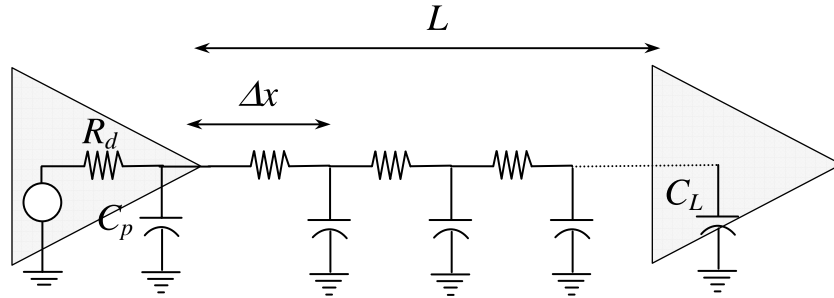

Figure 6. A distributed *RC* interconnect line model driven by source driving resistance R_d and terminated at load capacitance C_L.

The line is partitioned into n equal segments, each with length Δx. By using the distributed *RC* Elmore delay model, delay D for signal propagation through the line can be written as follows:

$$D = R_d((\sum_{i=1}^{n} c_0(x_i) \cdot \Delta x) + C_L) + \sum_{i=1}^{n} r_0(x_i) \cdot \Delta x \cdot (\sum_{j=i}^{n} c_0(x_j) \cdot \Delta x + C_L) \qquad (3.1)$$

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

$$D = R_d(C_L + \int_0^L c_0(x)dx) + \int_0^L r_0(x) \cdot (\int_x^L c_0(\tau)d\tau + C_L)dx \tag{3.2}$$

The third integral in (3.2) represents the downstream capacitance seen by the interconnect line from location x. It is assumed that the capacitance per unit length 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 R_d will be constant at the chosen operating temperature of the cell. By using (2.7), we can simplify (3.2) as follows:

$$D = D_0 + (c_0 L + C_L)\rho_0 \beta \int_0^L T(x)dx - c_0 \rho_0 \beta \int_0^L x \cdot T(x)dx \tag{3.3}$$

where:

$$D_0 = R_d(C_L + c_0 L) + (c_0 \rho_0 \frac{L^2}{2} + \rho_0 L C_L) \tag{3.4}$$

D_0 is the Elmore delay of the interconnect corresponding to the unit length resistance at reference temperature.

From (3.3) it is clear that in order to calculate the actual temperature-dependence, the area under $T(x)$ and $xT(x)$ should get evaluated first. To get an idea as to how much temperature can affect the degradation of the delay, we assume a worst case scenario by using a uniform thermal profile at some peak temperature over the entire length of the interconnect line. 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 (27°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=0.32 μm, R_d=10Ω, and C_L=1000fF for different lengths (in μm) are summarized in Figure 7. As Figure 7 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 interconnect length in (3.3). 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 line 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 length [20].

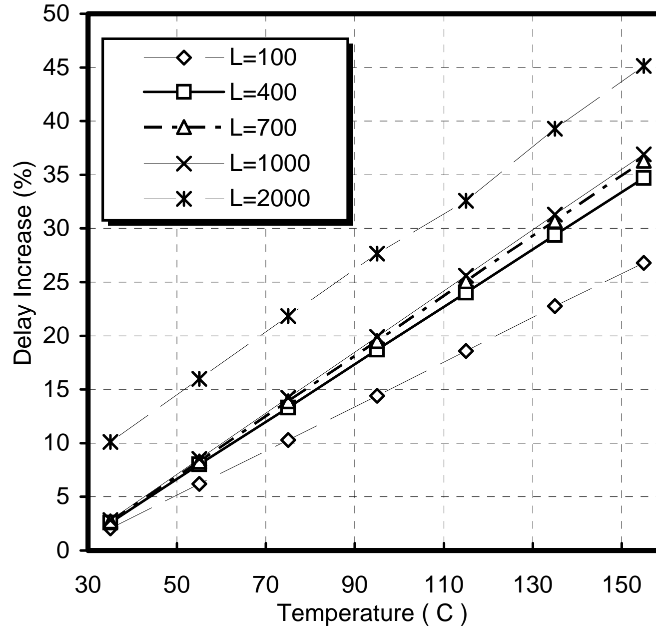

Figure 7. Percentage increase in signal delay with respect to nominal delay at room temperature (27ºC) as a function of the average interconnect line temperature for various lengths.

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 most of the solutions to the interconnect heat transfer equation (2.13) usually have an exponential component. By applying an exponential thermal distribution *T(x) = a.exp(-bx)* to an interconnect and using (3.3), the Elmore delay is as follows:

$$D = D_0 + \frac{a}{b}\rho_0 \beta[(c_0 L + C_L - \frac{c_0}{b}) + (\frac{c_0}{b} - C_L)e^{-bL}] \qquad (3.5)$$

where D_0 is defined by (3.4). For the sake of analysis, consider two different exponential thermal profiles $T_1(x)$ and $T_2(x)$ along an arbitrary interconnect line as depicted in Figure 8.

By using (3.3), calculation shows that the interconnect Elmore delay is more adversely affected by $T_1(x)$ than by $T_2(x)$, even though the underlying areas for both $T_1(x)$ and $T_2(x)$ in Figure 7 are equal along the length of the line. Figure 9 compares the performance degradation in the presence of $T_1(x)$ and $T_2(x)$ in two different wire lengths, 1000 μm and 2000 μm, relative to the case with a uniform line temperature at reference temperature (27 °C) having the same electro-thermal characteristics as mentioned before. In all scenarios the lower-bound temperature, T_L, is kept constant at 40 °C. By increasing the upper-bound value, T_H, for these functions it can be observed that using $T_2(x)$ causes less delay increase than that caused by using $T_1(x)$. 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 will be illustrated later in more detail. The above observation demonstrates that if we have the choice, choosing thermal profile $T_2(x)$ over $T_1(x)$ is preferable. Figure 9 also demonstrates that optimizing thermal gradients is as important as minimizing interconnect length for delay optimization purposes.

It must be noted that the substrate thermal map is strongly dependent on the design, synthesis, floor-planning 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 10.

By applying $T(x) = T_{\max} \cdot \exp(-(x-\mu)^2 / 2\sigma^2)$ to (3.3) the interconnect performance degradation can be examined. The movement of median μ along the length of the line will change the magnitude 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 $\mu = 0$ $(0 \leq x \leq L)$, 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 11 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}=110 °C and the same electrical and thermal properties as described above for Figure 7. 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.

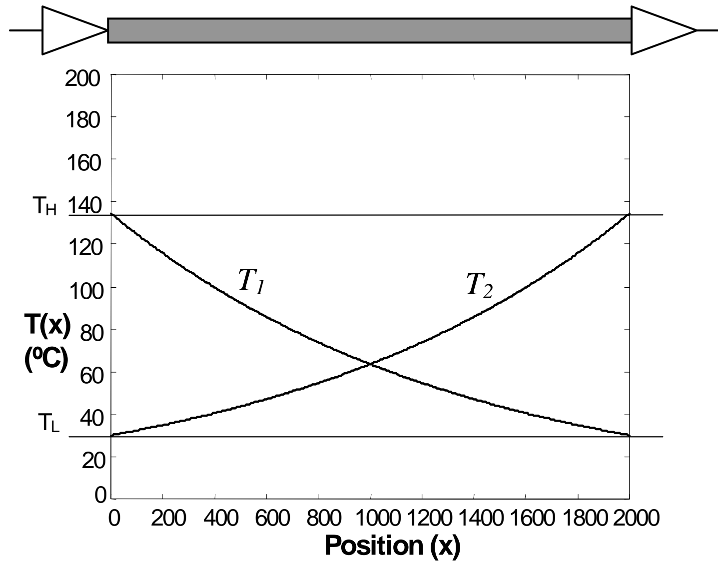

Figure 8. Schematic of two exponentially-distributed thermal profiles in different directions along the length of an interconnect line used to examine the effect of such non-uniformities on the signal propagation performance. In the worst case, both thermal profiles impose an excessive gradient of (+/-) 110ºC along the length of the interconnect line.

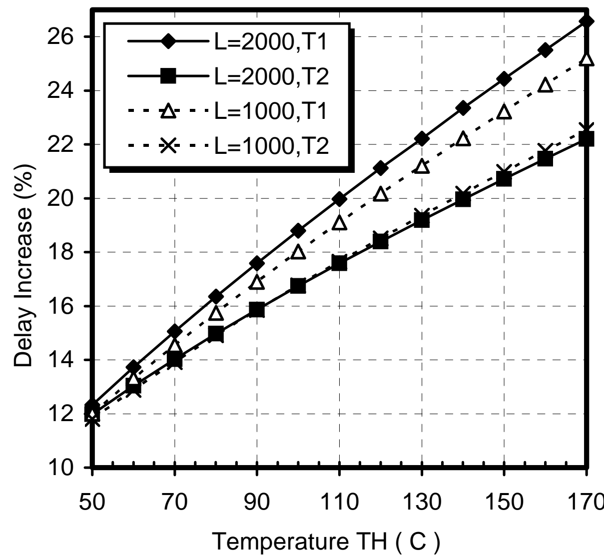

Figure 9. Performance degradation in a 2000 µm long interconnect line subject to thermal profiles T_1 and T_2 (c.f. Figure 8) as compared to the delay of the same line at uniform temperature of 27 °C. The x-axis shows the value of T_H in Figure 8, while T_L assumed to be fixed at 40 °C.

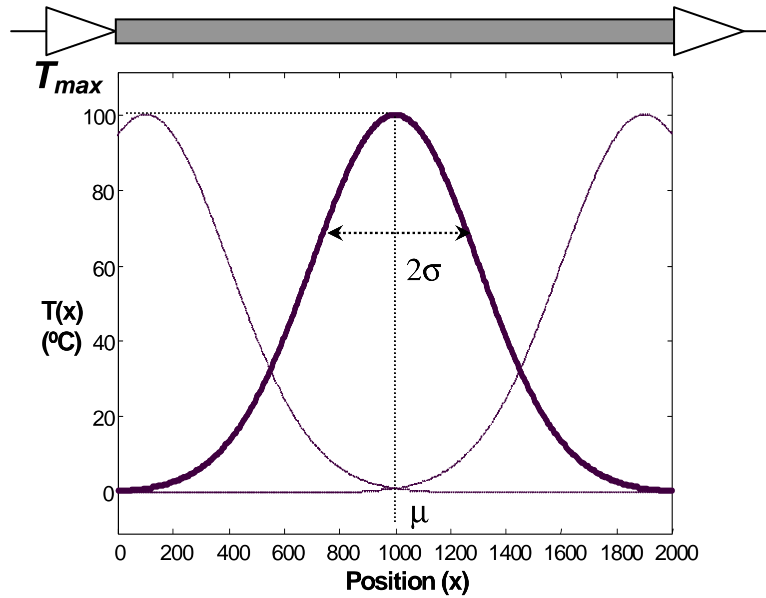

Figure 10. Constant-peak normally-distributed thermal profile with variable median μ and standard deviation σ along a 2000 µm long interconnect line.

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 [30], 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 an otherwise uniform resistance line, the results are therefore analogous to optimal uniform resistance non-uniform wire sizing (by assuming a constant capacitance).

It should be noted that with the current technology design methodologies and objectives in high-performance ICs, the worst-case temperature gradients observed on the substrate surface are around 50 °C in chips with substantially large area [16]. Use of large thermal gradients (per unit length) in previous examples (and also in next sections) is been done to simply highlight the possible impact of such gradients on the performance degradation in next generation ICs. This suggests that designers should also consider limiting the magnitude of such gradients in addition to other design constraints. Neglecting the

presence of excessive substrate thermal gradients resulting due to various non-uniform switching activity profiles as well as aggressive dynamic power management design techniques will consequently degrade the performance.

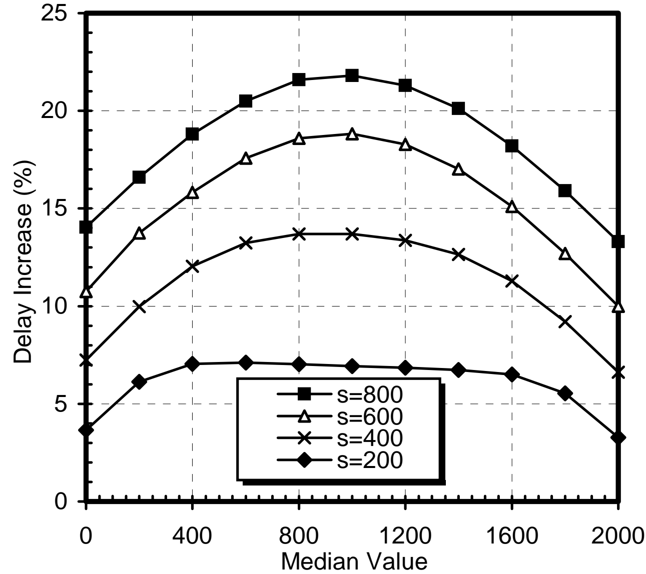

Figure 11. Delay degradation of a 2000 μm long interconnect line subject to a constant-peak normally-distributed thermal profile shown in Figure 10 as a function of its median value for various standard deviations (s), comparing to the delay of the same line at uniform room temperature of 27°C.

3.1 Error Estimation Based on Average Line Temperature

Even though (3.3) gives an accurate method of interconnect delay calculation based on the non-uniform line temperature, in practice the designers usually assign a pre-defined *constant* temperature to an interconnect line in the circuit and calculate the interconnect delay value based on a constant resistivity at that pre-defined temperature. The pre-defined temperature value is usually set to be at the maximum temperature corner of the cell library (T_{max} = 110~120 °C). Figure 12 depicts the percentage of error between the delay calculated by (3.3) for a 2000μm line due to non-uniform thermal profiles defined in Figure 8 and the Elmore delay calculated for the same line at a constant line temperature of 110°C. Like before, by increasing the upper bound temperature T_H (and fixing the lower bound T_L at 40 °C) one can determine the effect of the magnitude of interconnect thermal gradient on the outcome.

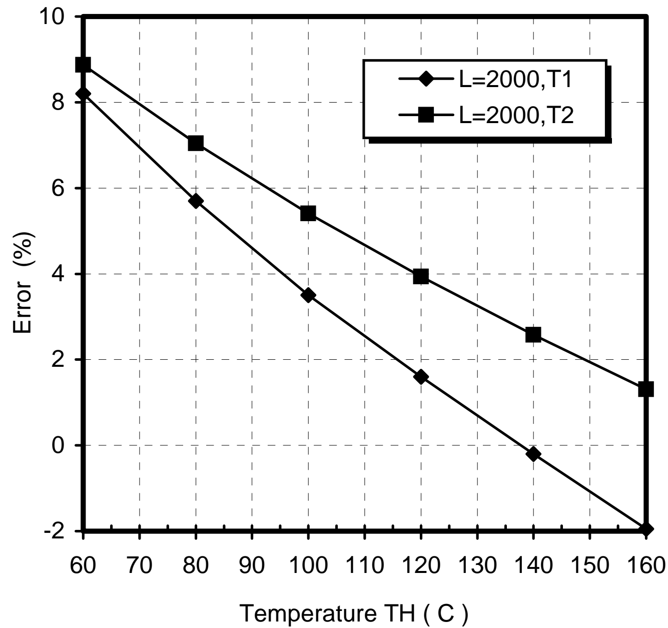

Figure 12. **Percentage of delay difference between the non-uniform temperature-dependent interconnect delay derived by (3.3) and the delay at a uniform line temperature T_{max}=110°C. Figure 8 depicts the non-uniform temperature profiles T_1 and T_2. The x-axis shows the value of T_H in Figure 8, while T_L assumed to be fixed at 40 °C.**

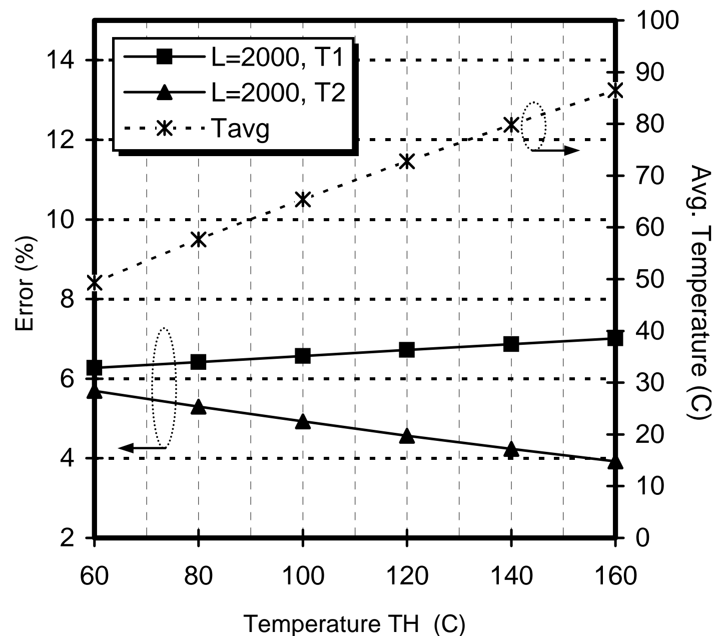

Figure 13. **Percentage of delay difference between the non-uniform temperature-dependent interconnect delay derived by (3.3) and the delay at a uniform line temperature T_{avg}. Figure 8 depicts the non-uniform temperature profiles T_1 and T_2. The x-axis shows the value of T_H in Figure 8, while T_L assumed to be fixed at 40 °C.**

It can be observed that by using the maximum substrate temperature, T_{max}, as the uniform line temperature, a significant amount of error get introduced into the interconnect delay calculation methodology. In this example, as the temperature of the upper bound increases, the difference between the delays estimated using the two methodologies decreases. This suggests that as the *average* temperature of the line gets closer to T_{max}, the error should decrease gradually. Intuitively, this generally means that using the average temperature instead of T_{max} should result in a better solution with less error (when compared to the actual non-uniform delay calculation based on (3.3)). Based on the thermal profile $T(x)$ along the length of an interconnect line, one can calculate the average temperature (T_{avg}) as follows:

$$T_{avg} = \frac{1}{L}\int_{T_0}^{T_L} T(x)dx \qquad (3.6)$$

where L is the length of the line and T_0 and T_L are the temperatures at the two ends of the line.

Figure 13 depicts the percentage error between the delay calculated by (3.3) for a 2000 μm line due to thermal profiles defined by Figure 8 and the Elmore delay calculated for the same line at a constant line temperature T_{avg}. It can be observed that even though there is still a certain amount of error between the two methods of delay calculation, using the T_{avg} results in less error, and its value is less dependent on the magnitude of the gradient (compared to Figure 12). Figure 14 depicts the percentage of delay difference between the delay calculation based on the non-uniform thermal profile defined by Figure 10 and the Elmore delay at uniform line temperature T_{max} (110°C). It is obvious that the amount of the delay difference is extremely dependent on the shape and magnitude of the gradient (in this example, depending on deviation σ and median μ). Even though in some special cases the magnitude of error is close to zero, considering the uniform line temperature T_{max} as the basis for Elmore delay calculation will result in an unpredictable amount of error. However, based on Figure 15, by using T_{avg} as the uniform line temperature, the error incurred in comparison to the one from actual non-uniform temperature dependent delay values is much less than that incurred by using T_{max}. In addition, the percentage of delay difference is almost independent of the shape of the interconnect non-uniform temperature (in this example the magnitude of error is between 5 to 7 percent regardless of median and deviation).

These examples show that in order to have a rough estimation of the interconnect delay based on a uniform line temperature, using T_{avg} as the line temperature is a much more accurate methodology than using T_{max} instead. However, the major shortcoming of this method is that the effects of the *direction* of

thermal gradient on the interconnect delay (which was discussed earlier) cannot be addressed by simply calculating the average line temperature T_{avg}. As shown later, the skew resulted by non-uniform temperature profiles in a clock tree is a function of directional thermal gradients and one should derive the interconnect delay based on the non-uniform temperature of each individual line in the clock tree instead of using a uniform line temperature.

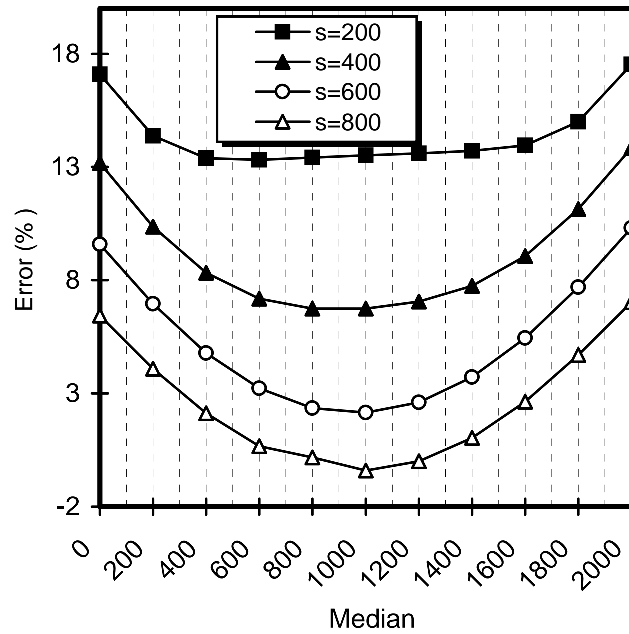

Figure 14. Percentage of delay difference between the non-uniform temperature-dependent interconnect delay derived by (3.3) and the delay of the same line at uniform temperature T_{max}=110°C as a function of median μ for different deviations (s). Figure 10 depicts the non-uniform temperature profile.

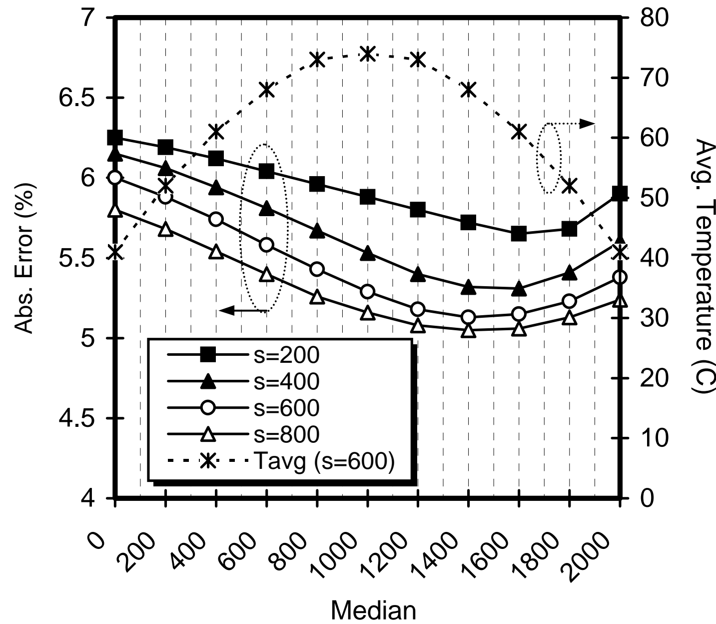

Figure 15. Absolute percentage of delay difference between the non-uniform temperature-dependent interconnect delay derived by (3.3) and the delay of the same line at a uniform temperature T_{avg} as a function of median μ for different deviation (s). Figure 10 depicts the non-uniform temperature profile.

4 Impact of Non-uniform Interconnect Temperature on Clock Skew

As shown in Section 3, the increase in the Elmore delay in global interconnects can be significant at high temperatures. Moreover, delay variations arising from non-uniform interconnect thermal profiles cannot be only accounted for by estimating a worst-case delay based on a uniform 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 further. Assume a clock net with two fanouts as illustrated in Figure 16. For simplicity assume that both wires 1 and 2 have the same lengths, widths, and electro-thermal characteristics (as used in Section 3) and are routed on the same layer.

Assuming different but uniform temperature profiles along both wires, the signal skew can be extracted from Figure 7 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. For the sake of analysis, lets 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 16 (for wire 2), with the length x at temperature T_2 and the length *(L-x)* at temperature T_3. Figure 17 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.

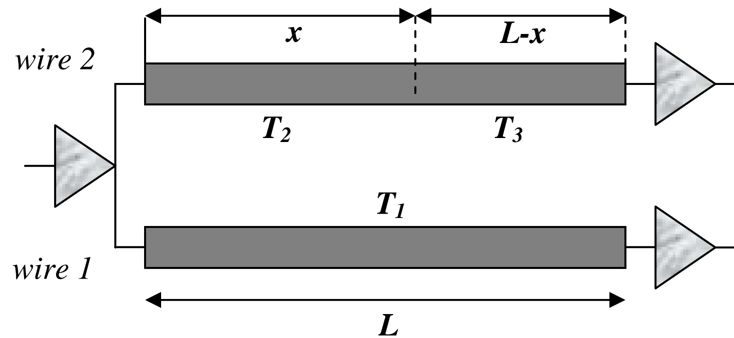

Figure 16. A simple portion of a clock tree with two fanouts and equal segment lengths.

Now assume that while wire 1 remains 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 18 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 significant thermal gradients is much higher than that for the shorter

signal nets. As a result, non-uniform substrate temperature profiles should be carefully considered during the clock network synthesis and routing stages [31].

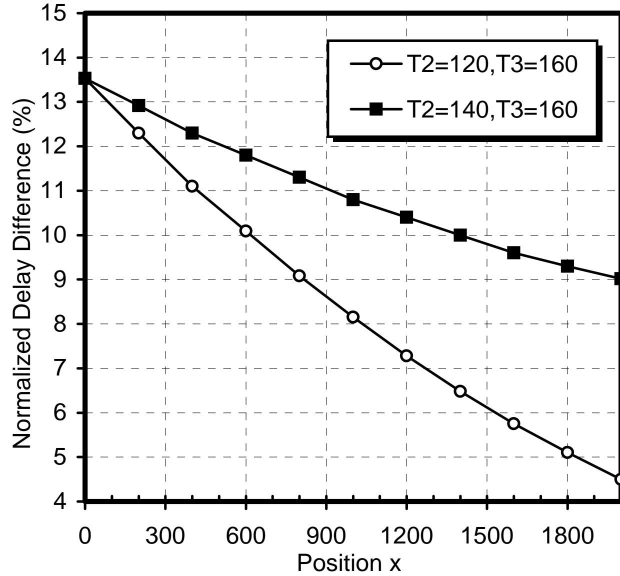

Figure 17. Percentage of normalized delay difference between wire segment 1 and wire segment 2 (c.f. Figure 16) as a function of break point *x*.

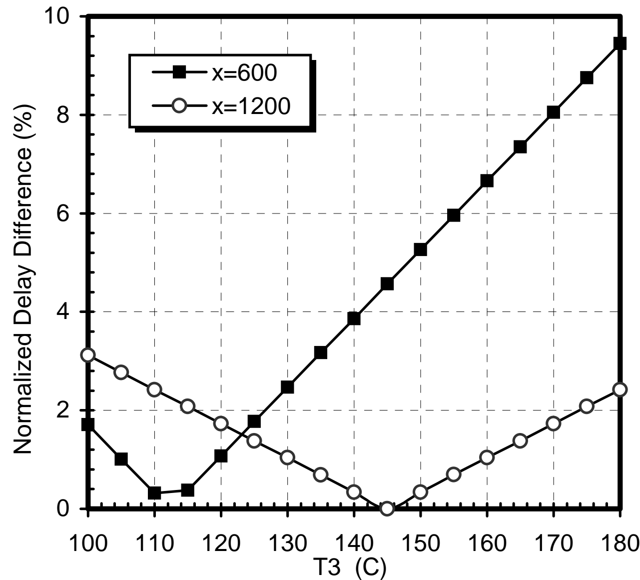

Figure 18. Percentage of normalized delay difference between wire segment 1 and wire segment 2 as a function of temperature T_3 as depicted in Figure 16, assuming T_1=100 ºC and T_2=80 ºC.

4.1 Temperature-Dependent Zero Skew Clock Routing Methodology

One of the basic goals of a typical clock signal distribution network is to maintain a zero (or near-zero) skew among the distributed clock to the sink elements. To ensure zero skew clock distribution, symmetric H-Tree structure or bottom-up merging technique may be used [32]. For simplicity and without loss of generality, for our analysis we consider the H-Tree clock topology consisting of trunks (vertical stripes) and branches (horizontal stripes) as depicted in Figure 19. In general, the top-level segments of the tree are wider than the lower level segments. Furthermore, the top-level global segments of the tree are assigned to the upper metal layers and low-level local segments are routed using the lower metal layers [33].

The problem arises due to the fact that trunk 1 and branches 2 of the H-Tree are sufficiently long. Hence, there is a high probability that those wire segments get exposed to the thermal non-uniformities in the underlying substrate. Such non-uniformity results in different signal delays at the two ends of trunk 1 and branches 2 of the H-Tree (Figure 19), hence there will be a non-zero skew among the sinks of the clock tree. Therefore, the non-uniform thermal effects result in a scenario where the symmetry of the H-tree cannot guarantee the zero skew among the sinks. If, for example, trunk 1 experiences a non-uniform thermal profile, the clock driver must be connected to this segment at a place other than the center of the segment. This also suggests that during a bottom-up construction of the clock tree, the actual temperature-dependent signal delay should be considered. Recently-reported thermal gradients of over 50 °C in some high performance designs [16], justifies the importance of this kind of analysis. Notice that the steady-state thermal profile of the substrate is being considered in this analysis. Even though the dynamic behavior of the chip causes transient changes in the cell switching activities, because of the large time constant for the temperature propagation in the substrate (around a few *ms*), the locations of the hot spots are in fact quite stable. It has been observed that the locations of most of the hot spots over the substrate are not strongly dependent on the application being executed. One major source of hot spots is the clock circuitry and its network components, and the power consumption in these hot spots is usually independent of the application being executed. Also, it has been shown that “usually”, specific blocks of high-performance microprocessors will become the hot spots regardless of the instruction flow, for example the instruction file [34].

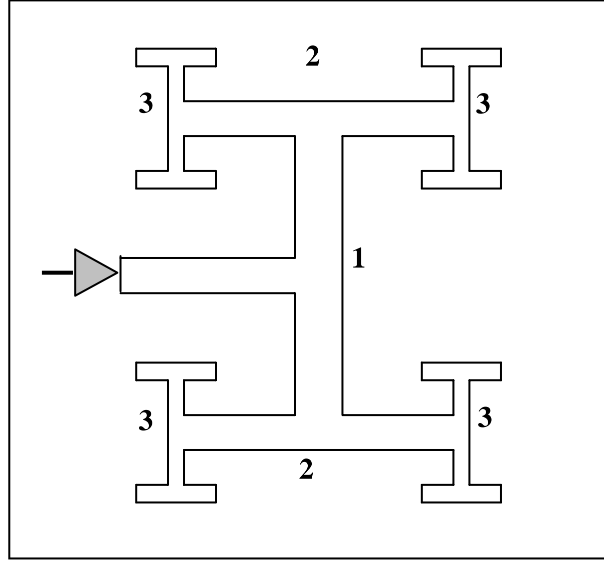

Figure 19. Illustration of a simple symmetric H-Tree clock distribution net, in order to satisfy zero-skew constraint at the clock sinks.

Consider the global trunk 1 in a portion of H-Tree magnified in Figure 20. The goal is to find the division point x along the length of the segment (L) such that when the clock signal driver is connected to that point, the delay at the two ends of the trunk 1 are the same. This will in turn ensure the minimal effect of non-uniform substrate temperature on clock skew. Assume an interconnect thermal profile $T(x)$ along the length L of trunk 1. By using the delay model described in Section 3, the thermally-dependent propagation delay from the source to the two ends of the trunk can be easily determined. By doing so and assuming balanced loads at the two ends p and q of the trunk and using (3.3), the optimum length l^* for ensuring zero clock skew can be obtained by solving the following equation:

$$\beta \int_0^{l^*} T(x)\,dx + l^* - A = 0 \qquad (4.1)$$

where A is a constant and it can be expanded as follows:

$$A = \frac{1}{Lc_0 + C_L}\left(\frac{L^2 c_0}{2} + LC_L + \beta(Lc_0 + C_L)\int_0^L T(x)\,dx - c_0 \beta \int_0^L xT(x)\,dx\right) \qquad (4.2)$$

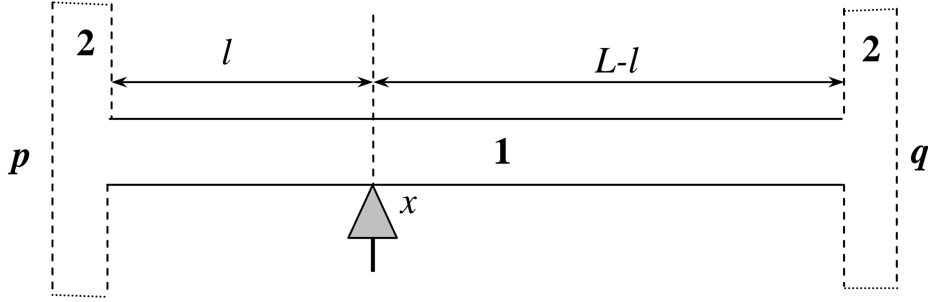

Figure 20. Schematic of minimum-skew clock signal insertion point for an interconnect line that is subjected to a non-uniform temperature profile along its length.

Given circuit parameters L, C_L, c_0, β and $T(x)$, one can easily compute constant A and solve (4.1) to obtain the optimal position for the clock signal connection to the net segment in order to minimize the clock skew at the two ends of the trunk. From (4.1) and (4.2), it can be seen that with a constant thermal profile $T(x)$ along the length of interconnect, connecting the clock signal at $l=L/2$ guarantees a zero skew. In fact, even a non-uniform, but symmetrical thermal profile with the symmetry axis at $l=L/2$ will result in a zero clock skew when the driver is connected to the middle of the line. From (4.1), it can also be seen that a gradually decreasing (increasing) thermal profile along the length of the line from 0 to L (from p to q), results in the optimum length l^* to be less than (greater than) $L/2$.

4.2 Experimental Results

By applying three different kind of interconnect thermal profiles, the behavior of temperature-dependent clock skew for a 2000 μm line with identical electro-thermal characteristics as those in Section 3 has been examined. The effects of linear, exponential and normal (Gaussian distribution with constant peak amplitude) thermal profiles on the clock skew were studied. Since the global clock lines are *thermally* long (i.e., longer than the heat diffusion length), we neglect the thermal effects of vias/contacts at the junction of the interconnect with the driver/receiver. In the first two cases, based on high temperature levels (T_H °C) and low temperature levels (T_L °C), different scenarios have been examined (Table 1). Column 3 shows the value of l^* at which, by inserting the clock signal in the H-Tree segment, zero clock skew at the two ends of the line is guaranteed. The reported normalized skew percentage in the fourth column represents the ratio of the clock skew when $l=L/2$ over the delay from the driver to any endpoint of the interconnect line when $l=l^*$. The third set of thermal profiles uses a constant-peak amplitude normal distribution with peak T_{max} (°C) at 100 °C, mean μ (μm) and standard deviation σ (μm), which imitates the behavior of a hot spot on the

substrate. Because this profile is symmetric, by applying a distribution with median *L/2*, a zero skew is guaranteed. Moving the hot spot along the length of the line clearly introduces the skew into the clock network.

From Table 1, it is clear that neglecting the effects of thermal profiles on the delay fluctuations, changes the worst-case clock skew behavior significantly. This suggests that for a given thermal profile *T(x)*, one should adjust the length of *l* by using (4.1) and (4.2) to maintain a zero clock skew. The circuit designer can place the cells such that the hot spots have a symmetrical position relative to the higher-level segments of the clock tree or can route the clock tree such that the higher level segments are symmetrical relative to the underlying hot spots. Because the number of these high-level clock segments is small, it is feasible to adjust the position of the clock tree segment or the cell placement over the substrate to maintain a nearly symmetric thermal profile along the clock segments.

Thermal Profile	***Parameters***	***l=l****	***Normalized Skew % L=L/2***
$T(x)=ax+b$, $a=\frac{T_H-T_L}{L}$, $b=T_L$	T_H=170, T_L=90	1042	5.42
	T_H=170, T_L=110	1032	3.98
	T_H=170, T_L=130	1021	2.65
	T_H=170, T_L=150	1012	1.29
$T(x)=a\cdot e^{-bx}$, $a=T_H$, $b=\frac{1}{L}\ln(\frac{T_H}{T_L})$	T_H=170, T_L=90	957.5	5.24
	T_H=170, T_L=110	968.66	3.63
	T_H=170, T_L=130	979.5	2.40
	T_H=170, T_L=150	989.7	1.19
$T(x)=T_{max}\cdot e^{-\frac{-(x-\mu)^2}{2\sigma^2}}$	μ=2000, σ=1000	1210	7.78
	μ=1000, σ=400	1000	0.0
	μ=500, σ=400	827	10.7
	μ=300, σ=700	911	9.57

Table 1. Impact of different thermal profiles on the skew degradation in a global trunk of an H-Tree based clock distribution network.

5 Conclusions

It was shown that non-uniform temperature distributions along global wires in high-performance ICs have significant implications for interconnect performance. 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 could be significantly impacted by the interconnect temperature non-uniformities. These studies reveal the necessity of incorporating the non-uniform chip thermal analysis during various optimization and planning steps in physical-synthesis flow in high performance IC designs.

Acknowledgement

The authors would like to thank Lukas P.P.P. van Ginneken for several helpful technical discussions during the initial phase of this work.

6 References

[1] V. De and S. Borkar, “Technology and design challenges for low power and high performance,” *in Proc. Int’l Symp. on Low Power Electronics and Design*, 1999, pp. 163 –168.

[2] P.P. Gelsinger, “Microprocessors for the new millennium: Challenges, opportunities, and new fronties,” *in Proc. Int’l Solid-State Circuits Conference*, 2000, pp. 22-25.

[3] K. Banerjee, S-C. Lin, A. Keshavarzi, S. Narendra and V. De, “A self-consistent junction temperature estimation methodology for nanometer scale ICs with implications for performance and thermal management,” *IEEE International Electron Devices Meeting*, pp. 887-890, 2003.

[4] Y-K Cheng, P. Raha, C-C Teng, E. Rosenbaum, and S. Kang, “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.

[5] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu, “On thermal effects in deep sub-micron VLSI interconnects,” in *Proc. Design Automation Conference,* 1999, pp. 885-891.

[6] S. Rzepka, K. Banerjee, E. Meusel, and C. Hu, "Characterization of self-heating in advanced VLSI interconnect lines based on thermal finite element simulation," *IEEE Trans. on Components, Packaging and Manufacturing Technology-A*, vol. 21, no. 3, pp. 406-411, 1998.

[7] J.R. Black, "Electromigration- A brief survey and some recent results," *IEEE Trans. on Electron Devices*, ed-16, pp.338- 347, 1969.

[8] H.A. Schafft, "Thermal analysis of electromigration test structures," *IEEE Trans. on Electron Device*, vol. ed-34, no.3, 1987, pp. 664-672.

[9] J. Tao, J.F. Chen, N.W. Cheung, C. Hu, "Electromigration design rules for bi-directional current", *in Proc. International Reliability Physics Symposium*, 1996, pp. 180-187.

[10] Y-K Cheng *et al*, "iCET: A complete chip-level thermal reliability diagnosis tool for CMOS VLSI chips," in *Proc. Design Automation Conference*, 1996, pp. 548-551.

[11] S. Im and K. Banerjee, "Full chip thermal analysis of planar (2-D) and vertically integrated (3-D) high performance ICs," *in Proc. Int'l Electron Device Meeting,* 2000, pp. 727-730.

[12] Y. Cheng, C. Tsai, C. Teng, and S. Kang, *Electrothermal analysis of VLSI systems*, Kluwer Academic Publishers, 2000.

[13] Z. Yu, D. Yergeau, R.W. Dutton, S. Nakagawa, N. Chang, S. Lin, and W. Xie, "Full chip thermal simulation," *in Proc. Int'l Symposium on Quality Electronic Design,* 2000, pp. 145-149.

[14] Q. Wu, Q. Qiu, and M. Pedram, "Dynamic power management of complex systems using generalized stochastic Petri nets," *in Proc. Design Automation Conference*, 2000, pp. 352-356.

[15] P.E. Gronowski, W.J. Bowhill, R.P. Preston, M.K. Gowan, and R.L. Allmon, "High performance microprocessor design," *IEEE Journal of Solid-State Circuits*, pp. 676-686, 1998.

[16] S. Borkar, T. Karnik, S. Narendra, J. Tschanz, A. Keshavarzi, and V. De, "Parameter variation and impact on circuits and microarchitecture," *in Proc. Design Automation Conference*, 2003, pp. 338-342. .

[17] C.H. Tsai and 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.

[18] C.H. Tsai and S.M. Kang, "Fast temperature calculation for transient electrothermal simulation by mixed frequency/time domain thermal model reduction," *in Proc. Design Automation Conference*, 2000, pp. 750-755.

[19] M.T. Bohr, "Interconnect scaling- the real limiter to high performance ULSI," *in Proc. Int'l Electron Device Meeting*, 1995, pp. 241-244.

[20] A.H. Ajami, K. Banerjee, M. Pedram, and L.P.P.P. van Ginneken, "Analysis of non-uniform temperature-dependent interconnect performance in high performance ICs," *in Proc. Design Automation Conference*, 2001, pp. 567-572.

[21] A.H. Ajami, K. Banerjee, and M. Pedram, "Analysis of substrate thermal gradient effects on optimal buffer insertion," *in Proc. International Conference on Computer-Aided Design,* 2001, pp. 44-48.

[22] A.J. Chapman, *Fundamentals of heat transfer*, 4th ed., New York, Mcmillan, 1984.

[23] A.A. Bilotti, "Static temperature distribution in IC chips with isothermal heat sources," *IEEE Trans. on Electron Device*, ed-21, no. 3, pp. 217-226, 1974.

[24] R.V. Andrews, "Solving conductive heat transfer problems with electrical-analogue shape factors," *in Chemical Engineering Progress*, vol. 51, no. 2, pp. 67-71, 1955.

[25] D. Chen, E. Li, E. Rosenbaum, and S.M. Kang, "Interconnect thermal modeling for accurate simulation of circuit timing and reliability," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, no. 2, pp. 197-205, 2000.

[26] M. Igeta, K. Banerjee, G. Wu, C. Hu, and A. Majumdar, "Thermal characteristics of submicron via's studied by scanning Joule expansion microscopy," *IEEE Electron Device Letters*, vol. 21, no. 5, pp. 224-226, 2000.

[27] National technology roadmap for semiconductors (NTRS), 1997.

[28] T-Y Chiang, K. Banerjee, and K. C. Saraswat, "Effect of via separation and low-k dielectric materials on the thermal characteristics of Cu interconnects," *in Technical Digest IEEE International Electron Devices Meeting*, 2000, pp. 261-264.

[29] International technology roadmap for semiconductors (ITRS), 2001.

[30] C-P. Chen, Y-P. Chen, and D.F. Wong, "Optimal wire-sizing formula under the Elmore delay model," *in Proc. Design Automated Conference*, 1996, pp. 487-490.

[31] A.H. Ajami, M. Pedram, and K. Banerjee, "Effects of non-uniform substrate temperature on the clock signal integrity in high performance designs," *in Proc. Custom Integrated Circuits Conference*, 2001, pp. 233-236.

[32] T.H. Chao, Y. C.Hsu, J.M. Ho, K.D. Boese, and A.B. Kahng, "Zero skew clock routing with minimum wirelength," *IEEE Transaction on Circuits and Systems-II*, vol. 39, no. 11, pp. 799-814, 1992.

[33] P. Zarkesh-Ha, T. Mule, and J.D. Meindl, "Characterization and modeling of clock skew with process variation," *in Proc. Custom Integrated Circuits Conference.*, 1999, pp. 441-444.

[34] K. Skadron, M.R. Stan, W. Huang, S. Velusmay, K. Sankaranarayanan, and D. Tarjan "Temperature-aware Microarchitecture," *in Proc. International Symposium on Computer Architecture*, 2003, pp. 2-13.