Effects of Non-uniform Substrate Temperature on the Clock Signal Integrity in High Performance Designs*

Amir H. Ajami, Massoud Pedram
Department of Electrical Engineering, Systems
University of Southern California, Los Angeles, CA 90089
{aajami, massoud}@zugros.usc.edu

Kaustav Banerjee
Center for Integrated Systems
Stanford University, Stanford, CA 94305
kaustav@cis.stanford.edu

Abstract- *This paper presents the analysis and modeling of the non-uniform substrate temperature in high performance ICs and its effect on the integrity of the clock signal. Using a novel non-uniform temperature-dependent distributed RC interconnect delay model, the behavior of clock skew in presence of the substrate thermal gradients is analyzed and some design guidelines are provided to ensure the integrity of the clock signal.*

I. Introduction

The ever-increasing demand for more complex ULSI circuits with aggressive performance is leading to higher power dissipation and increasing die and interconnect temperature. Management of thermally related issues is rapidly becoming one of the most challenging efforts in high performance chip design. At the circuit level, thermal problems have important implications for performance and reliability [1]. Furthermore, it has been recently reported that significant temperature gradients on the silicon substrate can occur due to different activity and/or different sleep modes of various functional blocks in high-performance microprocessor chips [2]. Dynamic power management (DPM) and functional block clock gating can be major sources of such thermal gradients over the substrate. In addition, as the technology feature size shrinks down, the global metal layers that carry the clock signal are getting closer to the substrate [3]. Hence, the effect of the non-uniform substrate temperature on the interconnect temperature profile becomes more critical. It is therefore essential to study and model the effects of non-uniform interconnect thermal profile caused by non-uniform temperature gradients over the substrate on the signal performance, i.e. signal delay and clock skew.

A systematic way of calculating the thermal profile of interconnects is described in Section II. In Section III a non-uniform temperature dependent signal delay model is introduced. Section IV examines the fluctuations of clock skew caused by non-uniform substrate thermal gradients and suggests design techniques to ensure the integrity of the clock signal. Finally, concluding remarks are made in Section V.

II. Non-uniform Chip Thermal Profile

The main sources of temperature generation in the chip are the switching activities of the cells over the substrate and the Joule heating of the interconnects due to the current passing through them. In a high performance design the substrate temperature can reach up to 120 °C, and Joule heating can contribute further to the overall temperature of an interconnect [3] [4].

*This work was supported in part by SRC under contract number 98-DJ-606.

Due to the presence of many heat generation sources in the substrate and the complicated boundary conditions (because of the convective nature of the heat transfer between the bottom side of the chip and the ambient), finding an analytical solution for the heat diffusion equation is non-trivial. As a result, much of the research work has focused on obtaining a solution by using numerical techniques, most notably the fast thermal analysis (FTA) method [5]. However, it must be mentioned that the accuracy of this kind of analysis depends on how accurately the power consumption of each cell (or macro-cell) over the substrate can be determined. In general, the thermal profile over the surface of the substrate depends on the power consumption of the cells and the distances between them.

A. Analytical Model for Interconnect Temperature Profile

Using appropriate boundary conditions, heat flow in interconnect can be obtained by solving the heat diffusion equation in the *3-D* space. In the steady state, by assuming that the four sidewalls and the top surface of the chip are thermally isolated (these are generally valid assumptions), the heat diffusion equation can be reduced to a *1-D* form as follows:

$$\frac{d^2T}{dx^2} = -\frac{Q}{k} \quad (1)$$

where Q is the volumetric heat generation rate inside the interconnect (W/m^3) and k_m is the thermal conductivity of the interconnect material (W/m°C) which is assumed to be constant. Consider an interconnect with length L, width w and thickness t_m that passes over the substrate with an insulator of thickness t_{ins} and thermal conductivity k_{ins} separating the two. The interconnect is connected to the substrate by vias/contacts at its two ends. The volumetric heat generation in the interconnect is computed by determining the rate of power generation due to the RMS current and the rate of heat loss due to the heat transfer between the interconnect and the substrate through the insulator. As a result, the heat flow equation (1) in an interconnect can be restated as follows [6]:

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

where λ and θ are constants given as follows:

$$\lambda^2 = \frac{1}{k_m}\left(\frac{k_{ins}}{t_m \cdot t_{ins}} - \frac{I_{rms}^2 \cdot \rho \cdot \beta}{w^2 t^2}\right) \quad (3)$$

$$\theta = \frac{I_{rms}^2 \cdot \rho}{w^2 \cdot t_m^2 \cdot k_m} \quad (4)$$

T_{line} is the interconnect temperature as a function of position along the length of the interconnect (which we will refer to as the interconnect thermal profile), T_{ref} is the underlying substrate temperature, ρ is the metal electrical resistivity at the

0-7803-6591-7/01/$10.00 © 2001 IEEE

reference temperature (0 °C), and β is the temperature coefficient of resistance in 1/°C (see (6)). In order to have a unique solution for (2), we need to provide two boundary conditions. Equation (2) shows the importance of the substrate temperature profile T_{ref} in determining the interconnect temperature. When considering short local wires, T_{ref} is usually assumed to be a constant. For long global interconnects, this is not a valid assumption since these lines span a large area of the substrate surface. Due to different switching activities of the cells in the substrate, a non-uniform temperature gradient which is created by the so-called *hot spots* over the substrate, is inevitable. As a result, determining the substrate thermal profile is crucial to the thermal analysis of the interconnects.

B. Effect of Layer Assignment on Interconnect Temperature

From (3) and (4) it can be deduced that the thermal profile along an interconnect is strongly dependent on the thickness of the underlying insulator t_{ins}. It is obvious that for a given technology, interconnects assigned to higher metal layers are farther from the substrate and as a result, the thermal resistances between these interconnects and the substrate are larger. Therefore, the higher metal layers that carry the clock signal (and have higher current density than the lower metal layers [1]) experience higher temperatures in comparison to the lower metal layers (recall that the interconnect can only exchange energy with the substrate and the top side of the chip is assumed to be thermally isolated).

As a simple example, consider an interconnect line of length L that is connected to the substrate using vias placed at its end points. For simplicity let us assume that the substrate has a uniform temperature of T_0. Using the two boundary conditions $T(x=0)=T_0$ and $T(x=L)=T_0$, the interconnect thermal profile is obtained by solving (2) as follows:

$$T(x)=\frac{\theta}{\lambda^2}(1-\frac{\sinh\lambda x+\sinh\lambda(L-x)}{\sinh\lambda L})+T_0 \quad (5)$$

This thermal profile has been depicted in Figure 1. Clearly, the peak value of the temperature is at $T_0+\theta/\lambda^2$. From (3) and (4), we can see that a larger t_{ins} results in a larger peak value for the thermal profile. Distance L_{td} in Figure 1, which is defined as the thermal diffusion length, is important because for a line shorter than this length, the line temperature does not reach its peak value ($L_{td} \cong$ 40-50 um for *AlCu* interconnects [6]). Hence, it can be deduced that the peak temperature rise will be maximum for the global wires which are thermally long (i.e. $L>>2*L_{td}$) and also have the maximum underlying insulator thickness, t_{ins}.

C. Effect of Insulating Material on Interconnect Temperature

From (3), it can be deduced that an insulator with larger thermal conductivity k_{ins} causes the resulting temperature in the interconnect to be lower. Physically, this can be explained as follows. Because the metal line is in general hotter than the substrate, the heat flows through the insulator toward the substrate. If the insulator is a better thermal conductor, then the rate of heat flow between the interconnect and the substrate will be larger. As a result, the effective temperature in the interconnect will become lower. Hence, proper selection of the insulator material between the interconnect and the substrate is important to manage the peak value of the interconnect thermal distribution.

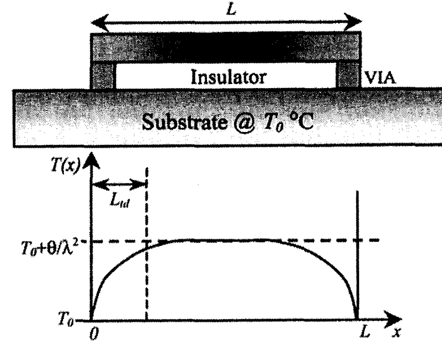

Figure 1: Thermal profile along a long interconnect line.

III. Non-uniform Temperature-Dependent Signal Delay Model

As mentioned before, the resistance of the interconnect has a linear relationship with its temperature and can be written as follows:

$$r(x)=r_0(1+\beta\cdot T(x)) \quad (6)$$

where r_0 is the unit length resistance at 0 °C and β is the temperature coefficient of resistance (1/°C). Consider an interconnect with length L and uniform width w that is driven by a driver of output resistance R_d and terminated by a load with capacitance C_L. The line is partitioned into n equal segments, each with length Δx. Using a distributed RC Elmore delay model and assuming that the number of parts n goes toward infinity, the delay D of a signal passing through the line can be written as follows:

$$D=R_d\cdot(C_L+\int_0^L c_0(x)dx)+\int_0^L r(x)\cdot(\int_x^L c_0(\tau)d\tau+C_L)dx \quad (7)$$

Assume that capacitance per unit length (c_0) does not change with temperature variations along the interconnect length (which is often a valid assumption) and also assume 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). By using (6), we can rewrite (7) as follows:

$$D=D_0+(c_0L+C_L)\,r_0\beta\int_0^L T(x)dx-c_0r_0\beta\int_0^L xT(x)dx \quad (8)$$

where:

$$D_0=R_d(C_L+c_0L)+(c_0 r_0\frac{L^2}{2}+r_0LC_L) \quad (9)$$

D_0 is the Elmore delay (at 0 °C) when the effect of temperature on the line resistance is neglected. Consider circuit parameters for *AlCu* interconnects with β=3E-03 (1/°C) and using r_{sh}=0.077(Ω/sq) at the reference room temperature (25 °C) and c_{sh}=0.268(fF/μm) as the unit sheet resistance and the unit length capacitance, respectively. In an interconnect with w=0.32 μm, R_d=10 Ω and C_L=1000 fF, for each 20 degree increase in the line temperature, there is roughly a 5 to 6 percent increase in the Elmore delay for a long global line (L>2000 μm). In this calculation we used a uniform thermal profile along the interconnect (the worst case scenario for delay degradation).

In reality, and especially for long global lines, the thermal profile along the length of an interconnect is non-uniform, as mentioned earlier. To understand the importance of considering the effect of non-uniform temperature on the delay, assume an interconnect whose two ends are at different temperatures and the profile between the two ends is modeled by an exponential distribution $T(x)=a.exp(-bx)$ with parameters a and b (Figure 2). We apply two different thermal profiles $T_1(x)$ and $T_2(x)$ along the length of the interconnect. Figure 3 compares the delay degradation in the presence of $T_1(x)$ and $T_2(x)$ in two different wire lengths, 1000 µm and 2000 µm, with identical electro-thermal characteristics as mentioned above. In both cases the lower bound temperature is kept constant at 30 °C. By increasing the upper bound value (x-axis of Figure 3) for these functions, it can be observed that using $T_2(x)$ causes less delay increase under the same conditions than when using $T_1(x)$. This shows that the assumption of a constant temperature along the wire (with peak-value) can introduce a large error in planning wire routings and clock-skew analysis. The above observation also demonstrates that if we have the choice, choosing the thermal profile $T_2(x)$ over $T_1(x)$ is preferable.

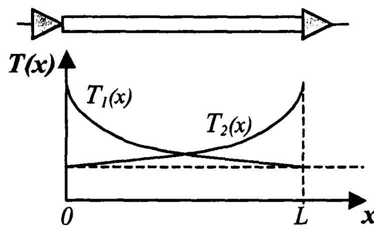

Figure 2: Schematic of exponential thermal profiles along an interconnect.

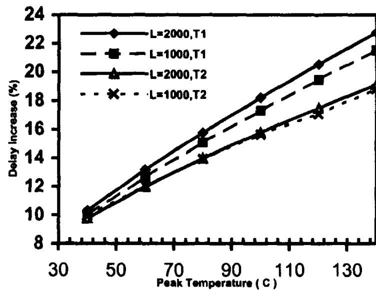

Figure 3: Degradation in interconnect performance caused by $T_1(x)$ and $T_2(x)$ (cf. Figure 2).

An explanation for the above behavior is that, from the resistance point of view, fluctuations of the temperature along the line are equivalent to wire sizing with uniform resistance. In sections with higher temperature, the wire can be modeled as a narrower wire and in sections with lower temperature the wire acts like a wider uniform resistance wire. As a result, an increasing thermal profile is equivalent to a decreasing sizing profile for a uniform resistance wire, which is known to give better delay than that with an increasing sizing profile [7].

IV. Effect of Non-uniform Substrate Temperature on Clock Skew

In addition to the performance degradation introduced by increasing temperature in the interconnect (which causes the effective signal delay to worsen), the non-uniform thermal profile along upper layer interconnects has a major effect on the skew of the clock signal net. The goal of the clock signal distribution network is to maintain a zero (or near-zero) skew through it. To ensure zero skew clock distribution, a symmetric H-Tree structure or a bottom-up merging technique can be used [8] [9]. 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 4. 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.

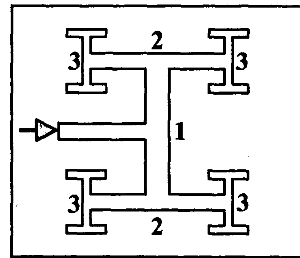

Figure 4: A symmetric H-Tree clock distribution net.

The problem arises from the fact that trunk 1 and branches 2 of the H-Tree are long. Hence, they are 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, hence there will be a non-zero skew along the tree. The temperature effects therefore result in a scenario where the H-tree symmetricity cannot guarantee the zero skew. 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 binary merge construction of the clock tree [8], the actual temperature-dependent delay must be considered. Having more than 30 °C thermal gradient in some designs [10], justifies the importance of this kind of analysis. Notice that we consider the *steady-state* thermal profile of the substrate. 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* [5]), the locations of the hot spots are in fact quite stable.

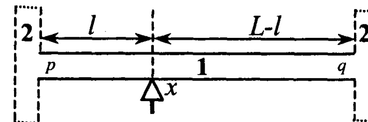

Figure 5: Schematic of minimum-skew clock signal insertion for an interconnect with non-uniform temperature profile.

Consider the global trunk 1 in the H-Tree depicted in Figure 5. 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 gradients temperature on skew. Assume an interconnect thermal profile *T(x)* along the length L of trunk 1 and by using the delay model described in III, we can write the propagation delay from the source to the two ends of the trunk. By doing so and assuming balanced loads at the two ends p and q of the trunk and using (8), the optimum length l^* for ensuring zero clock skew can be obtained by solving the following equation:

$$\beta \int_0^{l^*} T(x)\,dx + l^* - A = 0 \tag{10}$$

where A is a constant and can be written as follows:

$$A = \frac{1}{Lc_0 + C_L}\left(\frac{L^2 c_0}{2} + LC_L + \beta(Lc_0 + C_L)\int_0^L T(x)dx - c_0\beta\int_0^L xT(x)dx\right) \tag{11}$$

Given circuit parameters L, C_L, c_0, β and $T(x)$, we can easily compute the constant A and solve (10) to obtain the optimum position for the clock signal connection to the net segment. From (10) and (11), it is seen that with a constant thermal profile $T(x)$ along the length of interconnect, we can guarantee a zero skew by connecting the clock signal at $l=L/2$. 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 (10), we can also see 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) $L/2$.

We now examine the behavior of temperature-dependent clock skew for a 2000 μm line with identical electro-thermal characteristics as those in Section III, by applying three different interconnect thermal profiles. More precisely, we will consider the effects of linear, exponential and normal (Gaussian distribution with constant peak amplitude) thermal profiles on the clock skew. Since the global clock lines are thermally long, we neglect the thermal effects of vias/contacts at the junction of the interconnect with the driver/receiver. In the first two cases, different scenarios based on high temperature levels (T_H °C) and low temperature levels (T_L °C) have been examined (Table 1). Column 3 shows the value of l^* at which, by inserting the signal to the H-Tree segment, a zero clock skew is guaranteed. The reported normalized skew percentage in column 4 represents the ratio of the clock skew when $l=L/2$ over the delay from the driver to any endpoint of the interconnect 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 approximates the behavior of a hot spot on the substrate. As this profile is symmetric, by applying a distribution with median $L/2$, the zero skew is guaranteed. Moving the hot spot along the length of the line clearly increases the skew. It is clear from Table 1 that neglecting the effects of thermal profiles on the delay fluctuations, changes the skew by as much as 10 percent. The above discussion suggests that for a given thermal profile *T(x)*, one can adjust the length of l using (10) and (11) 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.

Table 1: Comparison between different thermal profiles and their effects on clock skew.

Thermal Profile	Parameters	*l**	*Normalized Skew %*
$T(x) = ax + b$, $a = \frac{T_H - T_L}{L}$, $b = T_L$	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

V. Conclusion

In conclusion, quantitative analysis of the interconnect thermal distributions arising from substrate thermal non-uniformities was presented. A detailed treatment of the impact of non-uniform temperature distributions on the interconnect performance was reported using a new distributed RC delay model that incorporates the non-uniform interconnect temperature dependency. It was shown that non-uniform temperature distributions along long global wires in high-performance ICs can have a significant impact on the interconnect performance and the worst-case clock skew. Finally, an analytical model that helps the designers in dealing with the non-uniformities in the interconnect thermal profile during the clock net routing has been presented for the first time.

References:

[1] K.Banerjee, A.Mehrotra, A.Sangiovanni-Vincentelli, and C.Hu, "On thermal effects in deep sub-micron VLSI interconnects," *36th ACM Design Automation Conference,*, 1999, pp. 885-891.

[2] Z.Yu, *et al.* "Full Chip Thermal Simulation," *Proceedings of the 1st IEEE Int. Symp. on Quality Electronic Design, March 20-22, 2000*, pp.145-149.

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

[4] 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.

[5] Y.Cheng, C.Tsai, C.Teng, S.Kang, "Electrothermal analysis of VLSI systems," *Kluwer Academic Publishers*, 1st ed., 2000.

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

[7] C-P.Chen, *et al.*, "Optimal wire-sizing formula under the Elmore delay model," *Proc. Design Automated Conference*, 1996, pp. 487-490.

[8] T.H.Chao, Y.C.Hsu, J.M.Ho, K.D.Boese, 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.

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

[10] P.E.Gronowski *et al.*, "High-performance microprocessor design," *IEEE Journal of Sold-State Circuits*, pp. 676-686, 1998.