An Efficient Method for Hot-spot Identification in ULSI Circuits

Yi-Kan Cheng†, and Sung-Mo Kang

†Somerset Design Center, Motorola Inc.
Semiconductor Products Sector
Austin, TX 78730
email: Yi-kan.Cheng@motorola.com

Coordinated Science Laboratory
Dept. of Electrical and Computer Engineering
Univ. of Illinois at Urbana-Champaign
Urbana, IL 61801

ABSTRACT

In this paper, we present a method to efficiently identify the on-chip hot spots in ULSI circuits. A set of mathematical formulae were derived in analytical forms so that local temperature information can be fetched quickly. These formulae were based on the Green's function and error function approximation, and the resulting equations were further simplified to a tractable level by asserting different constraints. Experimental result shows that this method is able to accurately locate the hot spots with little time complexity. It is particularly useful for temperature-driven circuit macro placement in early chip design phase, for which a large number of design iterations is needed and simulation efficiency is much required.

I. Introduction

Over the years, the state-of-the-art technologies have continued to push the ULSI chips to higher clock speed and packing density. The speed requirement causes large power consumption and high packing density results in large power density (power per unit area). One direct impact of the increasing power density is the dramatic on-chip temperature rise. For a chip under normal operating condition, the temperature rise can be as much as a few tens of degrees above the ambience. Without good thermal engineering, significantly non-uniform temperature distribution can lead to considerable on-chip temperature gradient. The temperature rise and temperature gradient has strong effects on both chip performance and reliability. Although many research efforts have been focusing on the development of low power and new package design for better IC reliability, thermal problems continue to grow and demand more attention.

Some research work has been done earlier to address the electrothermal problems in the device level and the small-scale integrated (SSI) level by developing the computer-aided design (CAD) tools [1][2]. The attempt at providing the electrothermal simulation capability at the VLSI level was introduced in ILLIADS-T [3]. Most of the existing electrothermal simulators focus on accurately calculating the temperatures at the expense of simulation time. In this paper, we present a new fast thermal analysis method, which has been incorporated into the thermal simulation framework in ILLIADS-T. It focuses on the on-chip hot-spot identification in an extremely efficient way. The hot spot information can be used to pinpoint the locations with high temperatures for reliability concerns. It can also be used to facilitate the adaptive mesh generation process for the numerical 3D thermal simulator in ILLIADS-T.

II. Thermal Analysis Framework

A. Background

The general heat diffusion equation for temperature calculation is [4]

$$\rho c_p \frac{\partial T(x,y,z,t)}{\partial t} = \nabla \cdot [k(x,y,z,T)\nabla T(x,y,z,t)] + g(x,y,z,t) \quad (1)$$

subject to the general thermal boundary condition:

$$k(x,y,z,T)\frac{\partial T(x,y,z,t)}{\partial n_i} + hT(x,y,z,t) = f_i(x,y,z). \quad (2)$$

In (1) and (2), T is temperature, g is power density of the heat source(s), k is thermal conductivity, ρ is density of material, c_p is specific heat, h is heat transfer coefficient between the object under simulation and the ambience, $f_i(x,y,z)$ is an arbitrary function, and n_i is the outward direction normal to the surface i. The details of the boundary condition specification can be found in [3].

B. Fast thermal analysis (FTA)

For a VLSI chip containing a large number of heat sources, the exact numerical method can be very expensive. In the early chip design phase, especially when no specific package information is given or the thermal boundary conditions are not fully characterized, a fast thermal analysis method emphasizing on the hot-spot identification is, therefore, much needed. We have developed a new fast thermal analysis (FTA) tool to meet such needs.

The FTA approach is based on the fact that, the dimensions of the gate-level or subcircuit-level heat sources in a VLSI chip are small compared with the chip size. Thus all heat sources can be viewed as located in a relatively *infinite* body. Consider a point source in a chip as shown in Fig. 1(a). Since ICs have a passivation layer, the top of the chip is insulated. We therefore have a boundary value problem with infinite dimension in the x-y plane while with semi-infinite dimension in the z direction. Moreover, the boundary condition (BC) at $z = 0$ is $\frac{\partial T(\mathbf{r},t)}{\partial z} = 0$. To find the temperature subject to this specific geometry and BCs, we use the *method of images* analogous to the electromagnetics problems [5]. We add an identical heat source symmetric

0-7803-5832-X /99/$10.00 ©1999 IEEE.

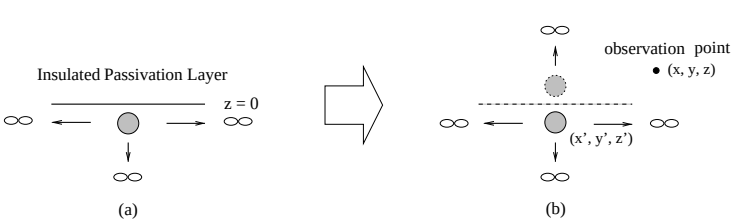

Figure 1: Method of images.

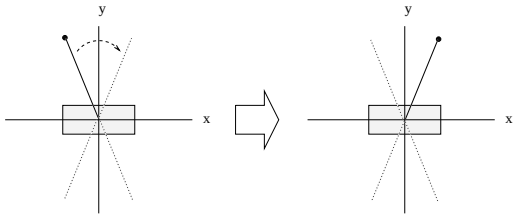

Figure 2: Transformation: constrain the observation point to the first quadrant.

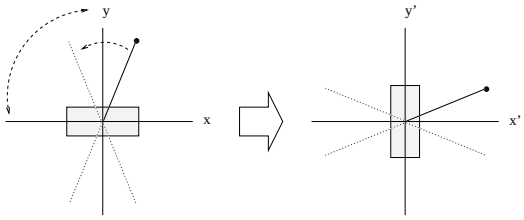

Figure 3: Transformation: constrain t_{a1} to be larger than t_{b1}.

with respect to $z = 0$ and remove the insulating boundary. Now the problem in Fig. 1(a) is transformed to that in Fig. 1(b).

The Green's function solution $G(\mathbf{r},t|\mathbf{r}',\tau)$ to the heat diffusion equation for the point source in Fig. 1 can be derived as $G(\mathbf{r},t|\mathbf{r}',\tau) = G_x \cdot G_y \cdot G_z$, where

$$\begin{aligned} G_x &= \frac{1}{[4\pi\alpha(t-\tau)]^{1/2}} \exp[-\frac{(x-x')^2}{4\alpha(t-\tau)}], \\ G_y &= \frac{1}{[4\pi\alpha(t-\tau)]^{1/2}} \exp[-\frac{(y-y')^2}{4\alpha(t-\tau)}], \\ G_z &= \frac{1}{[4\pi\alpha(t-\tau)]^{1/2}} [\exp(-\frac{(z-z')^2}{4\alpha(t-\tau)}) + \exp(-\frac{(z+z')^2}{4\alpha(t-\tau)})] \end{aligned} \quad (3)$$

where α is the thermal diffusivity. We formulate the resulting temperature rise above the ambience at observation point $\mathbf{r}$ due to parallelepiped heat source with dimensions $a \times b \times c$ as

$$\triangle T(\mathbf{r},t) = \frac{\alpha P_0}{k(abc)} \int_{\tau=0}^{t} \int_{-c}^{c} \int_{-b/2}^{b/2} \int_{-a/2}^{a/2} G(\mathbf{r},t|\mathbf{r}',\tau) dv' d\tau, \quad (4)$$

where the coordinate origin has been chosen to be at the center of the source and P_0 is the source power. To proceed, we integrate (4) by using error functions:

$$\triangle T(x,y,0,t) = \frac{\alpha P_0}{k(abc)} \int_0^t G(x,a,\tau) G(y,b,\tau) G(0,c,\tau) d\tau, \quad (5)$$

where the observation point is set to be on the chip surface ($z = 0$), and

$$\begin{aligned} G(x,a,\tau) &= \frac{1}{2}[\mathrm{erf}(\frac{a/2+x}{2\sqrt{\alpha(t-\tau)}}) + \mathrm{erf}(\frac{a/2-x}{2\sqrt{\alpha(t-\tau)}})], \\ G(y,b,\tau) &= \frac{1}{2}[\mathrm{erf}(\frac{b/2+y}{2\sqrt{\alpha(t-\tau)}}) + \mathrm{erf}(\frac{b/2-y}{2\sqrt{\alpha(t-\tau)}})], \text{ and} \\ G(0,c,\tau) &= \mathrm{erf}(\frac{c}{2\sqrt{\alpha(t-\tau)}}). \end{aligned} \quad (6)$$

Defining $\mathcal{A}_1 = 2(a/2+x)$, $\mathcal{A}_2 = 2(a/2-x)$, $\mathcal{B}_1 = 2(b/2+y)$, $\mathcal{B}_2 = 2(b/2-y)$, and $\mathcal{C} = 2c$, along with the change of variables, (5) can be recast as

$$\begin{aligned} \triangle T(x,y,0,t) = \frac{\alpha P_0}{4k(abc)} \int_0^t [\mathrm{erf}(\frac{\mathcal{A}_1}{4\sqrt{\alpha\tau}}) + \mathrm{erf}(\frac{\mathcal{A}_2}{4\sqrt{\alpha\tau}})] \cdot \\ [\mathrm{erf}(\frac{\mathcal{B}_1}{4\sqrt{\alpha\tau}}) + \mathrm{erf}(\frac{\mathcal{B}_2}{4\sqrt{\alpha\tau}})] \cdot \mathrm{erf}(\frac{\mathcal{C}}{4\sqrt{\alpha\tau}}) d\tau. \end{aligned} \quad (7)$$

In order to perform the integration in (7) analytically, we piecewisely linearize the error functions by [6]

$$\begin{aligned} \mathrm{erf}(x) &\approx 2x/\sqrt{\pi} \quad \text{for} \quad x \le \sqrt{\pi}/2; \\ &\approx 1 \quad \text{for} \quad x \ge \sqrt{\pi}/2. \end{aligned} \quad (8)$$

By defining $t_{a1} = \mathcal{A}_1^2/(4\pi\alpha)$, $t_{a2} = \mathcal{A}_2^2/(4\pi\alpha)$, $t_{b1} = \mathcal{B}_1^2/(4\pi\alpha)$, $t_{b2} = \mathcal{B}_2^2/(4\pi\alpha)$, and $t_c = \mathcal{C}^2/(4\pi\alpha)$, we have the approximations in Table 1, where $m = 0$ for $\mathcal{A}_2 > 0$ and $n = 0$ for $\mathcal{B}_2 > 0$, otherwise $m = 1$ for $\mathcal{A}_2 < 0$ and $n = 1$ for $\mathcal{B}_2 < 0$. To obtain the analytical solution of (7), we need to reduce the number of possible permutations (120) of t_{a1}, t_{a2}, t_{b1}, t_{b2} and t_c. To achieve this, we assert the following six constraints:

1. $x \ge 0, y \ge 0$
2. $(\frac{a}{2}+x) \ge |\frac{a}{2}-x| \Longrightarrow \mathcal{A}_1^2 \ge \mathcal{A}_2^2 \Longrightarrow t_{a1} \ge t_{a2}$
3. $(\frac{b}{2}+y) \ge |\frac{b}{2}-y| \Longrightarrow \mathcal{B}_1^2 \ge \mathcal{B}_2^2 \Longrightarrow t_{b1} \ge t_{b2}$
4. $2(\frac{a}{2}+x) \ge 2c \Longrightarrow \mathcal{A}_1^2 \ge \mathcal{C}^2 \Longrightarrow t_{a1} \ge t_c$
5. $2(\frac{b}{2}+y) \ge 2c \Longrightarrow \mathcal{B}_1^2 \ge \mathcal{C}^2 \Longrightarrow t_{b1} \ge t_c$
6. $t_{a1} \ge t_{b1}$

Constraints (2) and (3) are straightforward algebraically. Constraints (4) and (5) are valid because the thickness of heat sources is in the order of 0.1μm which is much smaller than the physical dimensions (i.e., $a \times b$) of a logic gate on chip.

Constraint (1) is equivalent to transforming all the observation points to the first quadrant by using the symmetric property as graphically shown in Fig. 2. To satisfy Constraint (6), we use the coordinate transformation as shown in Fig. 3. With the above specified constraints, (7) now becomes tractable. In other words, the precedence of t_{a1}, t_{a2}, t_{b1}, t_{b2}, and t_c must belong to one of the eight cases shown in Fig. 4. We have derived the analytical solutions for all cases. Proper solutions will be used during simulation, depending on the geometry and the size of the heat source, as well as the relative locations between the heat source and the observation point.

For a VLSI chip with n heat sources, the temperature rise at the center of source i is obtained by considering the heat diffusion from source i, plus that from other $n-1$ sources using superposition:

$$\triangle T_i^{\Sigma} = \triangle T_i(\mathbf{0},t) + \sum_{k=1}^{n-1} \triangle T_k(\mathbf{r}_k,t), \quad (9)$$

Table 1: Error function approximations.

$\mathrm{erf}(\frac{\mathcal{A}_1}{4\sqrt{\alpha\tau}})$	$\approx$	$(\frac{t_{a1}}{\tau})^{1/2}$	for	$\tau \geq t_{a1}$	$\mathrm{erf}(\frac{\mathcal{A}_2}{4\sqrt{\alpha\tau}})$	$\approx$	$(-1)^m(\frac{t_{a2}}{\tau})^{1/2}$	for	$\tau \geq t_{a2}$
	$\approx$	1	for	$\tau \leq t_{a1}$		$\approx$	$(-1)^m$	for	$\tau \leq t_{a2}$
$\mathrm{erf}(\frac{\mathcal{B}_1}{4\sqrt{\alpha\tau}})$	$\approx$	$(\frac{t_{b1}}{\tau})^{1/2}$	for	$\tau \geq t_{b1}$	$\mathrm{erf}(\frac{\mathcal{B}_2}{4\sqrt{\alpha\tau}})$	$\approx$	$(-1)^n(\frac{t_{b2}}{\tau})^{1/2}$	for	$\tau \geq t_{b2}$
	$\approx$	1	for	$\tau \leq t_{b1}$		$\approx$	$(-1)^n$	for	$\tau \leq t_{b2}$
$\mathrm{erf}(\frac{\mathcal{C}}{4\sqrt{\alpha\tau}})$	$\approx$	$(\frac{t_c}{\tau})^{1/2}$	for	$\tau \geq t_c$					
	$\approx$	1	for	$\tau \leq t_c$					

1. $t_{a1} \geq t_{b1} \geq t_{b2} \geq t_{a2} \geq t_c$
2. $t_{a1} \geq t_{b1} \geq t_{a2} \geq t_{b2} \geq t_c$
3. $t_{a1} \geq t_{a2} \geq t_{b1} \geq t_{b2} \geq t_c$
4. $t_{a1} \geq t_{a2} \geq t_{b1} \geq t_c \geq t_{b2}$
5. $t_{a1} \geq t_{b1} \geq t_{a2} \geq t_c \geq t_{b2}$
6. $t_{a1} \geq t_{b1} \geq t_{b2} \geq t_c \geq t_{a2}$
7. $t_{a1} \geq t_{b1} \geq t_c \geq t_{b2} \geq t_{a2}$
8. $t_{a1} \geq t_{b1} \geq t_c \geq t_{a2} \geq t_{b2}$

Figure 4: Eight cases under six constraints.

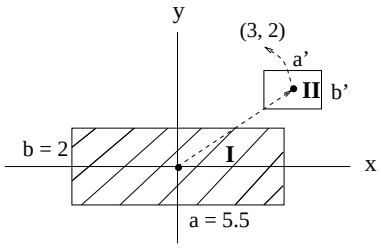

Figure 5: FTA example.

where $\triangle T_i^{\Sigma}$ is the temperature rise at the center of source i, $\triangle T_i(\mathbf{0},t)$ is the temperature rise due to i itself, and $\triangle T_k(\mathbf{r}_k,t)$ is the temperature rise due to source k. $\triangle T_i(\mathbf{0},t)$ and $\triangle T_k(\mathbf{r}_k,t)$ can both be found by combining (7) with one of the eight cases in Fig. 4. Take Fig. 5 as an example, where one source with power P_I is located at (0,0) while the other one with power P_{II} is at (3,2). To find the temperature rise at the center of source II due to source I (i.e., $\triangle T_k(\mathbf{r}_k,t)$ in (9)), we observe that it belongs to case 1 where $t_{a1} \geq t_{b1} \geq t_{b2} \geq t_{a2} \geq t_c$. Thus the integration in (7) at steady state ($t \to \infty$) can be calculated as

$$\triangle T_I(x,y,0,\infty) = \frac{\alpha P_I}{4k(abc)}(\sqrt{t_{b1}t_c} - \sqrt{t_{b2}t_c}) \cdot (4 + 2\log(\frac{\mathcal{A}_1}{\mathcal{B}_1})) - (\sqrt{t_{b2}t_c} + \sqrt{t_{a2}t_c}) \cdot 2\log(\frac{\mathcal{B}_1}{\mathcal{B}_2}) \quad \text{for} \quad m = 1 \text{ and } n = 1, \tag{10}$$

where $(x,y) = (3,2)$. Similarly, the temperature rise at the center of source II due to self heating (i.e., $\triangle T_i(\mathbf{0},t)$ in (9)) can be found since it is just a special case ($t_{a1} = t_{a2} \geq t_{b1} = t_{b2} \geq t_c$) of case 3:

$$\triangle T_{II}(\mathbf{0},\infty) = \frac{P_{II}}{\pi k a'}[2 + \log(\frac{a'}{b'}) - \frac{c}{b'}]. \tag{11}$$

Finally, the steady-state temperature rise at the center of source II can be obtained, according to (9), as

$$\triangle T_{II}^{\Sigma} = \triangle T_{II}(\mathbf{0},\infty) + \triangle T_I(3,2,0,\infty). \tag{12}$$

Mathematical formulation of the FTA method is based on the closed-form Green's function with the assumption of semi-infinite boundary condition. In practice, this assumption is not totally valid due to the existence of package and heat sink in a chip. Therefore, the temperature rise in (9) best represents the relative value instead of the absolute temperature rise. However, the FTA method provides a quick estimate of the temperature distribution. This is particularly useful when the number of heat sources is large or when a large number of repeated thermal simulations are needed. It is also useful when the chip package specification is not known in the early design phase.

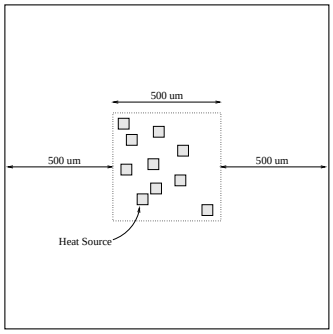

Figure 6: Chip structure and heat source locations.

III. Experimental Results

Among different thermal simulation methods, the FTA method is primarily used for fast hot-spot identification. In order to observe how accurately the FTA method can identify the hot spots, we performed the following experiment. Consider a chip containing 10 heat sources, all with dimensions of 50μm $\times$ 50μm. The sources are confined within the area (source area) with dimensions of 500μm $\times$ 500μm, and the distance between the boundary of the source area and the chip's bonding pad is 500μm as shown in Fig. 6. The heat sources were randomly placed and the power values from 10 mW to 100 mW were randomly assigned to the sources, and this process was repeated for 50 times (i.e., 50 tests). We claim that a violation happens in a test if the hot spot identified by the FTA method is different from the one identified by our numerical method developed earlier. The number of violation and the violation rate among 50 tests using the FTA method are shown in the second and third rows of Table 2 for different h (heat transfer coefficient in (2)) values. $\triangle T_{hot\_spot}$ represents the average difference of the actual temperatures of the two distinct hot spots identified by the FTA and numerical methods for all violated tests. The data shown in Table 2 reveal that a violation occurs only when the two hot spots have very close temperature values, which is insignificant for reliability concerns.

To demonstrate the advantage of FTA when the problem size

Table 2: Violation rate by using the FTA method.

$h\,[W/m^2 \cdot K]$	5,000	8,000	10,000	15,000	20,000	25,000	30,000	35,000
# violation	4	5	5	4	4	6	3	3
Violation rate (%)	8	10	10	8	8	12	6	6
$\triangle T_{hot\_spot}\ [^\circ C]$	1.1	0.83	0.6	0.8	0.35	0.75	0.21	0.16

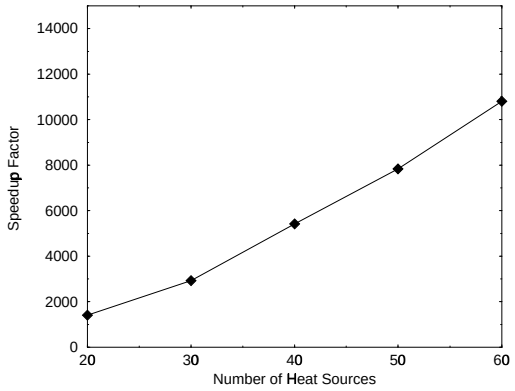

Figure 7: Speedup of FTA over numerical method.

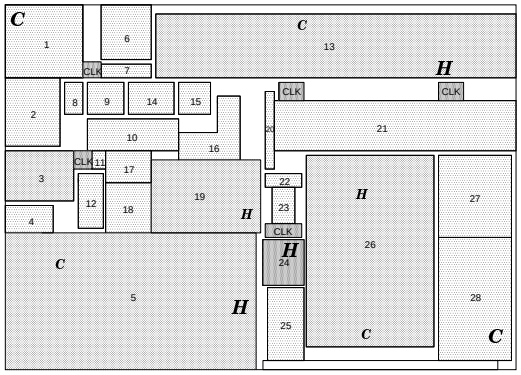

Figure 8: Unit-level layout of a prototype chip.

increases, both FTA and numerical method are used to identify the on-chip hot spots among all heat sources. The speedup factor of FTA over the numerical method with increasing number of heat sources is plotted in Fig. 7. In the numerical simulation, one grid line is assigned for each heat source in both x and y directions. This is to assign a minimum number of grids required to calculate the temperature of heat sources.

Another experiment is performed on a prototype chip with the unit-level layout shown in Fig. 8. Each unit contains several functional unit blocks (FUBs), and the power values were given for all FUBs. There are totally 510 FUBs in this chip. The symbols *H* and *C* in Fig. 8 represent the most significant hot and cold spots identified by FTA, while *H* and *C* represent the less significant hot and cold spots. The hot spots are caused primarily by placing two or more high power-density FUBs in close proximity. The result agrees very well with the detailed numerical simulation with package modeling [7], which is computationally expensive. By using FTA in the early chip design phase, it is possible to interchange the locations of hot and cold FUBs in order to ensure more uniform on-chip temperature profile and thus better reliability.

As a final remark, the efficiency of the iterative mesh generation technique employed by our numerical method depends on the initial guess of the number and locations of the mesh grids [3]. Chip area with larger temperature gradient often requires more grids for better accuracy. Currently the FTA routine has been incorporated into our numerical thermal simulator as a preprocessor to estimate the temperature gradient, which helps to determine a better initial guess of the grids and therefore facilitate a more accurate and efficient thermal simulation.

IV. Conclusion

In this paper, we presented a novel thermal analysis method - FTA. This method was developed to quickly find the on-chip hot spots for potential reliability concerns. FTA is based on the Green's function and error function approximation. With proper mathematical manipulation and constraint assertion, concise analytical formulae were derived for fast hot-spot identification. The FTA method is able to accurately locate the hot spots with comparatively little time complexity.

References

[1] K. Fukahori and P. R. Gray, "Computer simulation of integrated circuits in the presence of electrothermal interaction," *IEEE Journal of Solid-State Circuits*, vol. 11, pp. 834–846, Dec. 1976.

[2] C. H. Diaz, S. M. Kang, and C. Duvvury, "Circuit-level electrothermal simulation of electrical overstress failures in advanced MOS I/O protection devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 482–493, Apr. 1994.

[3] Y. K. Cheng, P. Raha, C. C. Teng, E. Rosenbaum, and S. M. Kang, "ILLIADS-T: An electrothermal simulator for thermal reliability diagnosis of CMOS VLSI chips," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 17, no. 8, pp. 668–681, Aug. 1998.

[4] M. N. Ozisik, *Boundary Value Problems of Heat Conduction*. New York, NY: Dover, 1968.

[5] C. A. Balanis, *Advanced Engineering Electromagnetics*. New York, NY: John Wiley & Sons, 1989.

[6] W. H. Beyer, *CRC Standard Mathematical Tables and Formulae*. Chemical Rubber Company, 1991.

[7] Y. K. Cheng, and S. M. Kang, "A temperature-sensitive simulation environment for thermally reliable ULSI chip design," submitted to *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Mar. 1999.