Architecture-Level Compact Thermal R-C Modeling

Marco Barcella, Wei Huang, Kevin Skadron, Mircea Stan

Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA 22904

Univ. of Virginia Dept. of Computer Science Tech. Report CS-2002-20, July 2002

Abstract

Architectural thermal modeling of high performance VLSI systems is of special importance in order to achieve reliable and power-saving system designs. In this paper, we present a novel, computation-efficient methodology for architectural level compact thermal modeling. The RC thermal models derived from this methodology are based on the layout of micro-architectural level functional blocks such as caches, execution units, and register files, etc. We utilize the duality between thermal and electrical circuit to perform electro-thermal simulations of both steady state and transient thermal responses of different blocks that dissipate different amounts of power. Temperature profiles derived from this modeling technique are compared with those derived from computation-intensive FEM/FDM tools. The model is both extremely fast and accurate, less than 4% error. This modeling methodology can then be adopted for architectural level VLSI design tools to perform fast electro-thermal simulations.

1. INTRODUCTION AND CONTRIBUTION

The advent of the SOC (system-on-chip) design style and the increasing power dissipation densities in VLSI systems make thermal awareness and modeling paramount. In high performance microprocessor design, the high cost of packages limits the architecture and circuit level solutions. The amount of heat that can be removed is limited. This is now one of the main obstacles to obtaining better performances in microprocessors. Moreover, each step in the design of an integrated circuit requires appropriate simulations. The dependency of the electrical simulations on the operating temperatures of the chip and the dependency of its temperature on its power dissipation require an electrothermal simulation to determine the operating conditions of the circuits. The electrical simulator, which computes the performance and tests the correct functioning of the device, also provides the power dissipations to the thermal solver that finds the temperature distribution on the chip. The temperature is used to update the models of the electrical circuits that are used by the electrical simulator. The iteration of these steps converges to an accurate temperature profile and circuit performance that takes into consideration the interdependence between the electrical and the thermal simulation.

The motivation for modeling heat at the architecture level is that architecture-level studies of thermal management typically don't have data or models for lower levels of abstraction. Moreover, an architecture-level analysis can be more useful than lower levels of abstraction, because the gains in power saving, thermal management and performance are typically higher.

The objective of this analysis is to provide fast thermal models that are accurate enough to make architectural level simulations of a VLSI system. In this paper, we consider the integrated circuits divided into blocks that are specified by the user. The power dissipation of each block is input into the model and its average temperature is determined. These models can be integrated in the Wattch and SimpleScalar framework, for example, creating a pipeline-level cycle-accurate electro-thermal simulator of a microprocessor. In this case, the blocks would correspond to the architectural components such as the cache and the branch predictor. A programmable way of producing compact RC models is provided. The model is made of thermal resistors, R's and thermal capacitors, C's. Testing these models entails finding an accurate solution of the temperature distribution that is used as a reference. A figure of merit based on the mean error is adopted. The reference solution can be found either with an analytical approach or with a numerical simulation. Both approaches have been explored, as the analytical one is very useful for simple components and simple boundary conditions while the numerical one can be used for the whole chip. We realized that a numerical simulation was actually complete and

suitable to the task and we adopted FloWorks as our numerical simulation tool (Finite Element Method)[1]. FloWorks is a software package provided by the europian Dassault Systems Company. The tool is available under an educational license. In our simulations in FloWorks, we use the function of heat transfer in solids to get our reference temperatures of different blocks.

These models need to be fast and accurate enough and capture the implications of a layout design. Results are valid for different chip layouts and for different power densities and power distributions. We also consider a simplified version of the model and a simplistic one, in which only vertical thermal resistances and capacitances are included.

2. RELATED WORK

Prior work has been done on thermal modeling of VLSI systems ([2]-[14]). The active research fields on this topic include thermal management, analytical and numerical approaches to thermal modeling, CAD tools and simulators for thermal modeling and management, and lumped thermal R-C networks. To the best of our knowledge, none of the work done before has reached a thermal model suitable for an electro-thermal simulation at the architectural level of a VLSI system. We develop a programmable thermal model that is based on the layout specification of the system and can be integrated in Wattch or SimpleScalar to generate a fast electro-thermal simulator.

In [2], the authors describe the motivations for thermal management for systems on chip (SOC) designs. Thermal management basically consists in monitoring the temperature on the chip and controlling the power dissipation so that the temperature targets and limits are respected. This allows using less expensive packages as less heat need to be removed to meet the temperature target. It entails the use of thermal simulation, modeling and temperature sensors. In [3], a software thermal simulator, MONSTR is implemented. The thermal modeling approach in [3] is based on the combination of numerical (FEM and FDM) and analytical (Fourier series) techniques. This solution method decomposes the original problem into a series of sub-problems. Hence, one can obtain analytical solutions at the lowest level and then combine these solutions into a general one by using the numerical (FEM) approach. Overall, this approach is accurate but still computationally intensive. A visualized thermal simulation tool, 2D-SUNRED, is presented in [4]. The authors used a finite difference method, FDM and an algorithm is used for successive node reduction, leading to a reduced thermal impedance matrix of the boundary nodes. The approach is not efficient enough for a thermal solver to be integrated in an architectural level electrothermal simulator.

In [5], the authors develop an analytical thermal model for on-chip heat dissipation in VLSI design. The boundary conditions are prescribed by a Neumann problem (adiabatic) on five faces of the die, and a Dirichlet (isothermal) problem on the bottom face (the one at contact with the heat-sink). In [6], another analytical thermal model is given. The boundary conditions here are: constant heat source on the top surface, adiabatic lateral sides, thermal conduction on inter-layer contacts and convective condition on bottom surface. Another analytical approach is discussed in [7]. In [8], [9], [10] and [11], electro-thermal simulation is considered. Numerical simulation of the temperature distribution is first considered. Both FEM (Finite Element Method) and FDM (Finite Difference Method) are computationally intensive and they can be used to evaluate simpler R-C models. They may also be used to generate faster models, by using reduction techniques and by increasing the grid spacing of the mesh. In [9], a rational formulation of the models is presented. In [12] an analysis in the frequency domain is used to determine the values of the R's and the C's in a R-C ladder thermal model. In [13], heat sink and packaging modeling is considered. In [14], a spreadsheet software package based method of generated R-C thermal networks is presented.

3. SET-UP

A chip composed of a silicon wafer divided into functional blocks, a heat spreader and a heat-sink made of 28 rectangular fins is modeled, Figure 1.The functional blocks corresponds to architecture-level structures like caches, branch predictor, instruction window … The model consists of a steady part, which is made of thermal resistors and a transient part made of thermal capacitors. The resistors and the capacitors together make the model. The model is programmed in C and is tested by using a reference solution. The reference solution is found by simulating both the heat transfer in the solids and the airflow around the heat sink. This is done using FloWorks that is a finite element numerical simulator. Since it is a tested numerical simulator that models the airflow as well as the heat flow in the solids, it is suitable for the analysis of a chip with its heat sink and it provides a very accurate reference. The average temperatures of each block are compared and the accuracy of the model is asserted.

The chip is considered as a representative of the current microprocessors: it is 10mm*10mm, it has a variable thickness, from 0.3mm to 0.7mm. Two heat sinks are considered, one in copper with a 0.78 K/W thermal resistance and one in aluminum with a 0.88 K/W thermal resistance. The model is tested for different thickness, heat sinks as well as different chip layouts, that is, different block configurations. It is also important that the model does not depend on the power densities of each block, as the user provides the power dissipation. Therefore, different power distributions were tested.

Three models are considered: the first one is the complete model, the second one is a simplified version and the last one is a simplistic version that does not use lateral thermal resistors but only vertical ones. The simplified model, instead, has lateral resistances but their values are simplified as shown in the next section. The results are analyzed separately for the steady state and the transient simulations in the next sections.

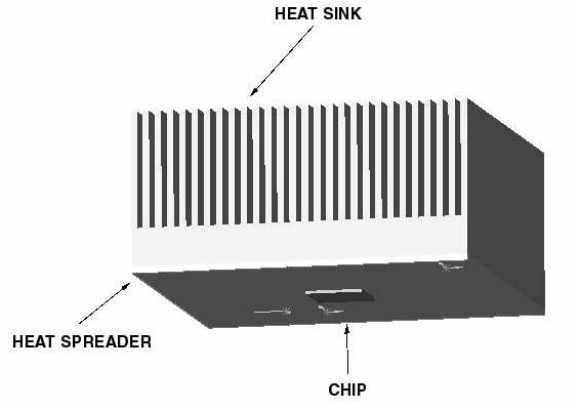


Figure 1. The system simulated: the air is forced to flow through the fins at 5 m/s.

4. COMPACT R-C MODEL (STEADY-STATE PART)

The compact R-C thermal model receives the representation of the chip layout with the power dissipation of each block and determines their average temperature. The steady-state thermal model (which uses only thermal resistors and no capacitors) of a sample chip layout with four blocks is shown in Figure 2.

This model is very fast because it is very small and the thermal resistances can be computed with minimum considerations of the neighboring blocks, i.e. the length and position of the shared edge. More details are given later in this section. The model can be determined and simulated in programming language.

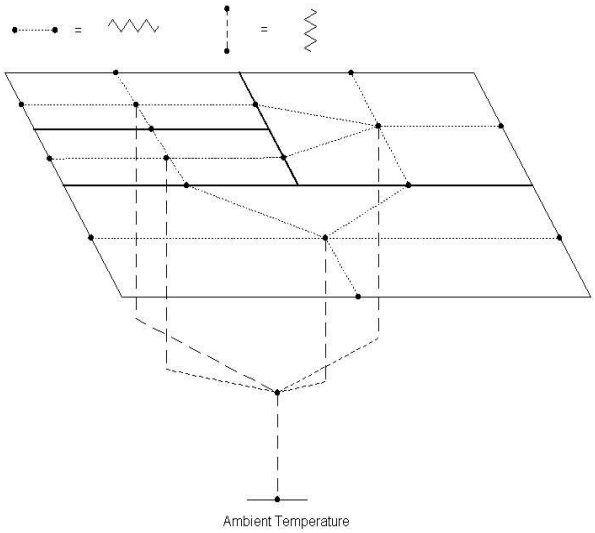


Figure 2. Example of steady-state model of a four-block layout.

The idea is to consider the layout as opposed to dividing the system in grids as in the finite element-meshing algorithm. Being tailored to the layout, the model can be smaller and faster and it can be integrated in SimpleScalar and Wattch as a very efficient thermal solver of an electro-thermal simulator. In this section, we only discuss our thermal network for steady state temperatures of different functional blocks. Therefore, only thermal resistances are considered. Transient models are discussed in the next section.

Since the vertical thermal resistances dominate the heat transfer in silicon chips, it is straightforward to think about a simplistic model without lateral thermal R's. But, the lateral R's do have effects on heat transfer. Therefore, we compared the simplistic model, which doesn't contain lateral thermal R's with our model that has lateral R's. The results show that the model with lateral R's provides much better results than the simpler model, as shown later in the paper. This is due to the lateral heat transfer in the silicon and mostly in the heat spreader.

As it can be seen in Figure 2, lumped thermal resistors connect the center of each block to the middle points of the shared edges with neighboring blocks. The values of the thermal resistances were found by considering a programmable systematic procedure that creates thermal R's from each of the blocks to all its neighbors. There are also thermal R's modeling the vertical heat transfer of the silicon wafer and the heat-spreader in each block. Finally a thermal resistance models the heat sink. These resistances

depend on both the physical parameters and the geometry of the chip. In particular, the ones between one specific block and one of its neighbors depend on the area of that specific block, its aspect ratio and the angle from the center of the specific block and the center of the shared edge of the neighboring block.

The procedure used to calculate the thermal resistances is thought as follows: given an arbitrary block with width W and length L, we compute the lateral thermal resistances R_x and R_y, as in Figure 3. This calculation for R_x and R_y is based on numerical simulations in FloWorks and is explained later in this section. Then, taking Figure 3 as an example, we consider the center of a block and we find the middle point of each shared edge of its neighboring blocks, for example l_1, l_2, and l_3 in Figure 3. Then we consider the R's from the center of the block to the middle points of that block's shared edges, shown as R_{31}, R_{32}, R_{21}, R_{23}, R_{12} and R_{13}. It can be seen that R_{31}, R_{32}, R_{21} and R_{23} are actually their corresponding R_{3x}, R_{3y}, R_{2y} and R_{2x}. For R_{12} and R_{13}, we can recognize that R_{12} and R_{13} are in parallel, and consider that the equivalent thermal resistance is R_{1x}, which is the horizontal thermal resistance of Block 1. Therefore, if we write the reciprocals of R_{12}, R_{13} and R_{1x} as G_{12}, G_{13} and G_{1x} we have:

$$G_{12} + G_{13} = G_{1x} \qquad (4.1)$$

or in a more general form,

$$G_X = \sum_{i=1}^{n} G_i \qquad (4.2)$$

Where, n is the number of neighbors of a specific block.

In order to solve G_{12} and G_{13} (supposing G_{1x} is known for now), we need another equation for G_{12} and G_{13} in addition to (2.1). In Figure 4, we draw the projection of the shared edge. The projection is perpendicular to the straight line connecting the center and the middle point of the edge. We did this by considering that the heat propagates to the neighboring block along the connecting straight line. Therefore, the effective heat transfer edge between two blocks is the projection of the shared edge, the cross-section of the ideal triangle.

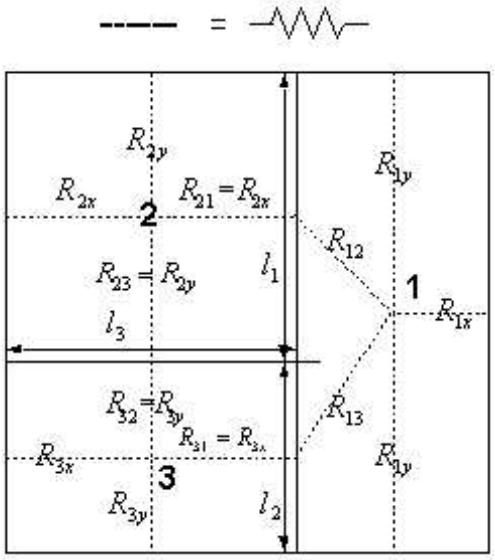


Figure 3. Calculation of lateral resistances (1)

We also considered that the thermal conductance is inversely proportional to the length of the connecting line and directly proportional to projected edge:

$$G_{12} \propto \frac{k \cdot l_1 \cdot \cos\theta_1}{L_1 / \cos\theta_1} = \frac{k \cdot l_1 \cos^2\theta_1}{L_1} \qquad (4.3)$$

And

$$G_{13} \propto \frac{k \cdot l_2 \cdot \cos\theta_2}{L_1 / \cos\theta_2} = \frac{k \cdot l_2 \cos^2\theta_2}{L_1} \qquad (4.4)$$

Therefore

$$\therefore \frac{G_{12}}{G_{13}} = \frac{l_1}{l_2} \cdot \frac{\cos^2\theta_1}{\cos^2\theta_2} \qquad (4.5)$$

Now, we have two equations, (2.1) and (2.5), with two unknowns. Thus G_{12} and G_{13} can be calculated. A Simplified Model is also considered where $G_{12}/G_{13}=l_1/l_2$, instead of (2.5).

R_x And R_y of each block are found considering some numerical simulation data (FloWorks). They are found to depend on both the area of the block and its aspect ratio, referring to Figure 5.

$$R_0 = \sqrt{W \cdot L} \cdot r_0 \qquad (4.6)$$

$$R_x = [b \cdot (\sqrt{\frac{L}{W}} - 1) + 1] \cdot \sqrt{W \cdot L} \cdot r_0 \qquad (4.7)$$

$$R_y = [b \cdot (\sqrt{\frac{W}{L}} - 1) + 1] \cdot \sqrt{W \cdot L} \cdot r_0 \qquad (4.8)$$

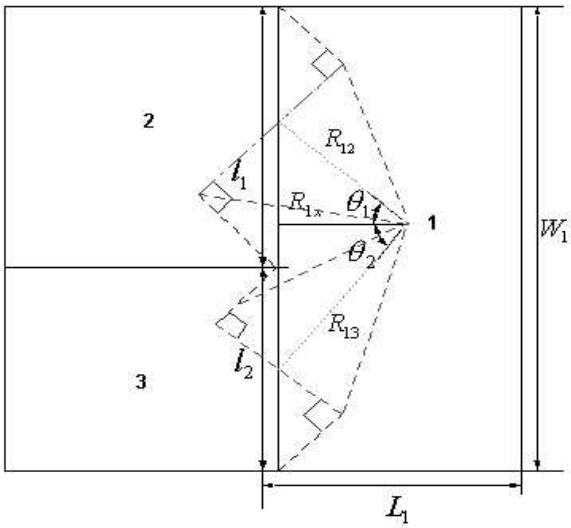


Figure 4. Calculation of lateral resistance (2)

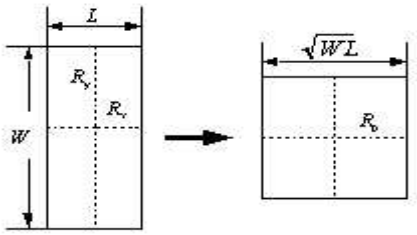


Figure 5. Calculation of the horizontal and vertical resistances (3**)**

In these equations, b ($b = 0.1166$) that models the aspect ratio dependence of the lateral thermal resistance. The coefficient b is constant. We considered the two relative lengths the above defined as $\frac{L}{\sqrt{L \cdot W}} = \sqrt{\frac{L}{W}}$ and $\frac{W}{\sqrt{L \cdot W}} = \sqrt{\frac{W}{L}}$ referring to Figure 5. ($\sqrt{L \cdot W}$ is the edge of the square with the same area of the original rectangular block) The ratios of R_x and R_y with R_o can be modeled as

$$\frac{R_x}{R_0} = b \cdot (\sqrt{\frac{L}{W}} - 1) + 1$$

And

$$\frac{R_y}{R_0} = b \cdot (\sqrt{\frac{W}{L}} - 1) + 1,$$

as shown from the FloWorks data in Figure 6. From these, (4.7) and (4.8) are found. R_0 is found to be proportional to $\sqrt{L \cdot W}$.

In (4.6), r_0 is the thermal resistance of a unit surface area:

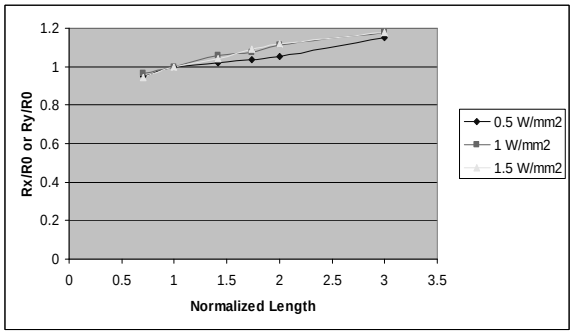


Figure 6. The Aspect Ratio dependence of the lateral resistances for different power densities.

$$r_0 = \rho \times \frac{length}{Area_{equivalent}} = \rho \times \frac{1}{thickness} \times \alpha \quad (4.9)$$

Here, ρ is the thermal resistivity of silicon, which is 10^{-2} m-K/W. We found that r_0 is inversely proportional to the thickness of chip, hence α is a constant dimensionless fitting parameter that is determined once for ever, after the first simulation. Different thickness, power distributions and chip layouts were considered as shown in the result sections. It was found that very good results could be found by always considering $\alpha = 1.8 * 10^{-4}$. This shows that the lateral thermal resistances are much smaller than expected considering just the silicon wafer, in fact they are multiplied by a α that is smaller than unity. This is mainly due to the extra lateral heat flow in heat spreader that effectively adds a resistance in parallel. In fact the heat flows laterally in parallel in the silicon wafer, in the heat spreader and in the heat sink. Moreover, the resistivity of copper and aluminum is about 4 times smaller than silicon and the thickness of the metal layers are about 10 times larger. Furthermore the heat flows continuously and not in lumped resistors. An additional study and modeling effort is underway. Alpha only multiplies the lateral resistances as shown.

Another task is to model the vertical thermal resistance of silicon chip, heat spreader and heat sink. In our model, each block has a separate thermal resistor for silicon as can be seen in Figure 7, the thermal resistance for heat spreader and heat sink were divided according to a factor s, in order to reach an isothermal surface located somewhere between the bottom surface of the silicon chip and the air interface of the heat sink. In fact, the heat spreader that we modeled, 1.00 mm of Copper proved to leave temperature differences on its lower surface, despite the fact that it is usually assumed to be isothermal. The total package resistance, R_{pack} =

$R_{spreader}$ + $R_{heatsink,}$ is divided into a part that is in common with all the blocks, R_{pack_com} = (1-s)*R_{pack} and the rest of the resistance is given by the resistances of the individual blocks, s*R_{pack}. We found that s = 120*t_k/$R_{pack,}$ tk being the thickness of the chip. For a 0.5 mm thickness, this means that 0.06 K/W of the heat-sink thermal resistance are modeled separately for each block as the isothermal surface is slightly lower than heat spreader and 0.72 K/W. The formula for s proved to be valid for both the heat sinks considered: R_{pack} = 0.78 K/W and R_{pack} = 0.88 K/W. The parameter is proportional to the thickness of the chip, because the silicon is the main responsible for the heat not spreading, so the more the silicon is thick, the more the heat doesn't spread. The heat-sink and heat spreader thermal resistance is hence divided into a part that is in common with each block and the remaining part that is divided in parallel resistances for each block that connect the lower surface of the silicon underneath each block and the resistance that is in common for every block.

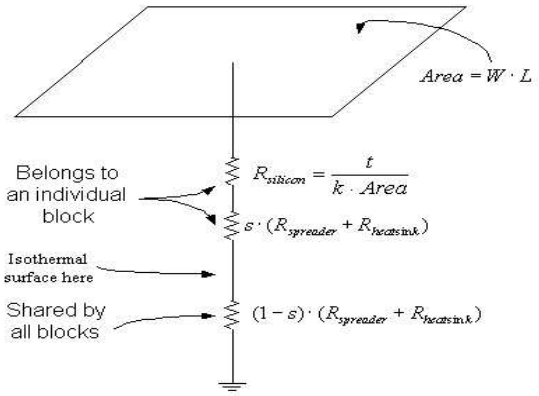


Figure 7. Vertical thermal resistances

5. COMPACT R-C MODEL (TRANSIENT PART)

Our proposed transient thermal model, which consists of adding thermal capacitances to the thermal resistances, is shown in Figure 8. The thermal capacitance is computed as

$$C = c \cdot volume \qquad (5.1)$$

Where c is specific heat. From (4.10), we can see that, since the equivalent thermal capacitance for two adjacent volumes is actually the sum of the two individual capacitances, the capacitances are modeled in parallel, no matter whether they are physically in parallel or in series with respect to the heat flow.

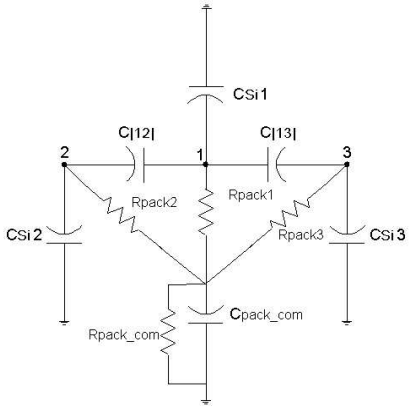


Figure 8. Transient model for the 3-block layout of Figure 4 (thermal resistances are the same as computed before, node 1 and 3 are not adjacent)

With this in mind, we model one combined capacitance between every two neighboring blocks, and model all the vertical capacitances to thermal ground. For lateral capacitances, we use the following formula:

$$C_{ij} = c \cdot \frac{1}{4} \cdot t \cdot W_i \cdot l_{ij} \qquad (5.2)$$

Where C_{ij} is the capacitance connecting the center of Block i to the edge shared with Block j, and c is specific heat, t is the thickness of the chip (we considered silicon thickness from 0.3mm to 0.7mm in the simulations presented here), W_i is the width of Block i, and l_{ij} is the shared length between block i and j. As it can be seen from the formula, the volume we considered here is the volume of a triangular cylinder with base area $\frac{1}{4} \cdot W_i \cdot l_{ij}$ and the chip thickness as height. According to equation (4.11), the total thermal capacitance $C_{|ij|}$ between Block i and Block j is:

$$C_{|ij|} = C_{ij} + C_{ji} = c \cdot \frac{1}{4} \cdot t \cdot (W_i + W_j) \cdot l_{ij} \qquad (5.3)$$

For the silicon vertical thermal capacitance of each block, we use the following formula:

$$C_{Si-i} = c \cdot t \cdot W_i \cdot L_i \qquad (5.4)$$

Putting these vertical capacitances in parallel with the thermal capacitance of the heat spreader and the heat sink, and adding all the thermal resistances calculated in the steady-state section, we finally find our compact transient thermal R-C model.

The effect of the isothermal surface on the partition of the package capacitance ($C_{package}$=300

J/K) into a component common to each block, C_{pack_com} and a component separate for each block, C_{Si}. As a first model, we consider only a common component that is defined as

$$C_{pack\_com} = C_{package} \cdot \frac{R_{package}}{R_{pack\_com}},$$

Where $R_{package}$ is the total package thermal resistance (0.78 K/W) and R_{pack_com} (0.72 K/W for a thickness of 0.5mm). This was done with the idea of preserving the main time constant:

$$R_{package} \cdot C_{package} = R_{pack\_com} \cdot C_{pack\_com}$$

This choice needs further investigation. Furthermore, the time constant is of the order of hundreds of seconds. For this reason, in the actual simulations done so far, $C_{package}$=1 J/K. This is done to have reasonable simulation times with FloWorks on our equipment.

The resulting model is shown in Figure 9, where layout 1 is considered as an example.

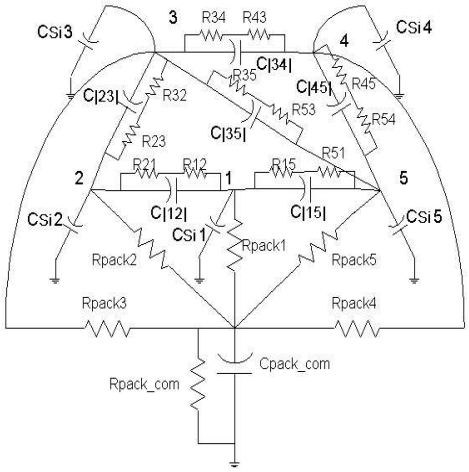


Figure 9. The model for Layout #2.

	Block 1	Block 2	Block 3	Block 4	Block 5
Block 1	—	R12=2K/W C|12|=0.005J/K	R13=36K/W C|13|=0.001J/K	—	R15=2K/W C|15|=0.0045J/K
Block 2	R21=1.3K/W C|12|=0.005J/K	—	R23=1.2K/W C|23|=0.00375J/K	—	—
Block 3	R31=8.4K/W C|13|=0.001J/K	R32=2.3K/W C|23|=0.00375J/K	—	R34=2.3K/W C|34|=0.005J/K	R35=7.2K/W C|35|=0.00075J/K
Block 4	—	—	R43=1.8K/W C|34|=0.005	—	R45=1.7K/W C|45|=0.0045
Block 5	R51=1.2K/W C|15|=0.0045J/K	—	R53=1.3K/W C|35|=0.00075J/K	R54=1.2K/W C|45|=0.0045J/K	—
	C1=0.014J/K	C2=0.006J/K	C3=0.012J/K	C4=0.012J/K	C5=0.006J/K
	Rpack1=0.9K/W	Rpack2=2.1K/W	Rpack3=1.05K/W	Rpack4=1.05K/W	Rpack5=2.1K/W
	Rpack_com=0.5K/W (shared by all blocks)				
	Cpack_com=1.4J/K				

Table 1. Values for the model in Figure 9.

6. SIMULATIONS AND EVALUATION (STEADY STATE)

We performed finite-element-method (FEM) simulations in FloWorks, and used the results as reference to test the accuracy of our steady state models. The two chip layouts are represented in Figure 10 (layout 1), and Figure 11 (layout 2). The first one is a case of 5 blocks that is used to prove the accuracy of the modeling approach. Several power distributions have been considered in order to prove the model with different total chip powers, different power densities and different block sizes, chip thickness and configurations. Table 2 shows the uniform block power dissipations assumed in the chip and the resulting power densities for the first layout considered (Figure 10). Table 2 shows the values for the second layout (Figure 11).

The base configuration that is considered is a chip 10mm×10mm with thickness 0.5 mm, layout configuration 1. Four results are compared, the first is the reference from FloWorks, the second is the complete model, the third is its simplified version, as specified in the previous section and the fourth is the simplistic one. The results are shown in Figure 12: both the model and its simplified version are very accurate, the simplistic model, instead does not contain the necessary information to model the transversal heat flow from one block to its neighbors.

Different power distributions are also considered so that the model is proven to work for each power input. One is shown in Figure 13. Another layout with 10 blocks, layout 2 is considered to prove that the model is independent on the layout and that it works even for a larger number of blocks, that is, a higher granularity, Figure 14.

The chip thickness is a parameter of the model (t_k) and so far we have only considered 0.5 mm, therefore, we show the results for 0.3 mm and 0.7 mm in Figure 15 and Figure 16, respectively.

The heat sinks are accurately designed and simulated in FLoWorks with a forced airflow of 5 m/s and 28 rectangular fins. In the base configuration, it is made of copper and the thermal resistance is 0.78 K/W. For Figure 17, the heat sink is considered to be in aluminum and its thermal resistance is 0.88 K/W. These real heat sinks are also representative of high performance microprocessors heat sinks, as shown in an article reported by Intel on heat sinks [15]. The results show that the model works the same for different heat sinks. We also considered extreme values of total power dissipation of 100 W and maximum power densities of 4 W/mm^2, that can be reached in the register file of a super-scalar processor, for example. As mentioned

before, this simplistic model does not contain any lateral resistances as if the blocks were transferring the heat directly to the heat sink independently, without exchanging it laterally between each other.

The comparison with FloWorks shows that both the complete model and its simplified version are very accurate, despite the fact that it is extremely compact and fast. The percentage error, computed as:

$$\frac{T_{FloWorks} - T_{Model}}{T_{FloWorks} - T_{Ambient}} \times 100\% \qquad (6.1)$$

Is always lower than 4%. The simplistic model (without lateral resistances) instead, is not accurate, showing that neighboring blocks do interact with each other.

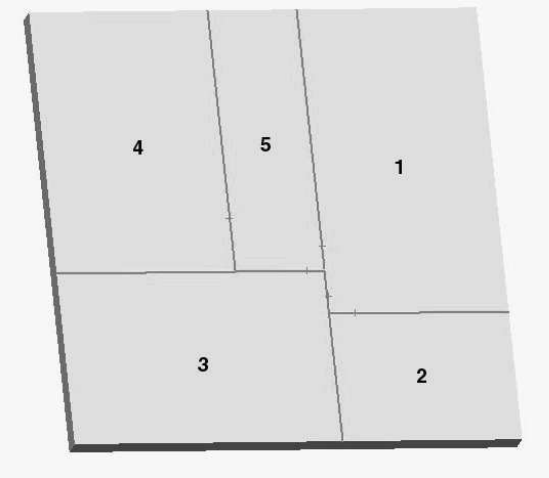


Figure 10. Layout #1

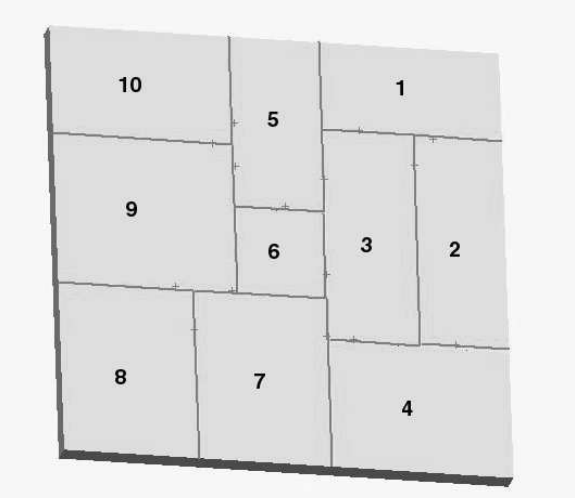
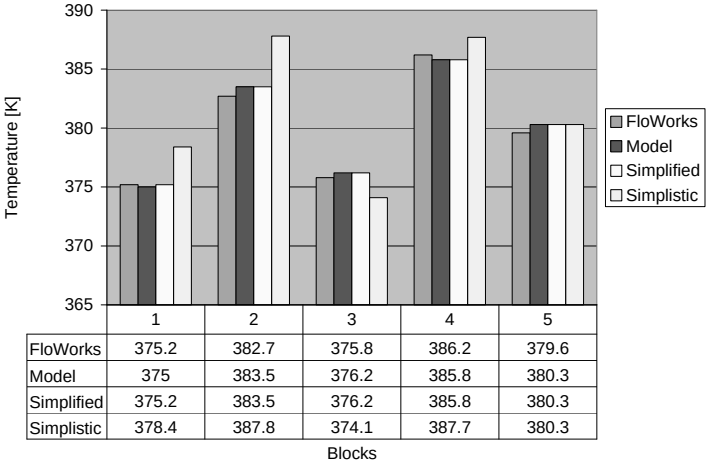
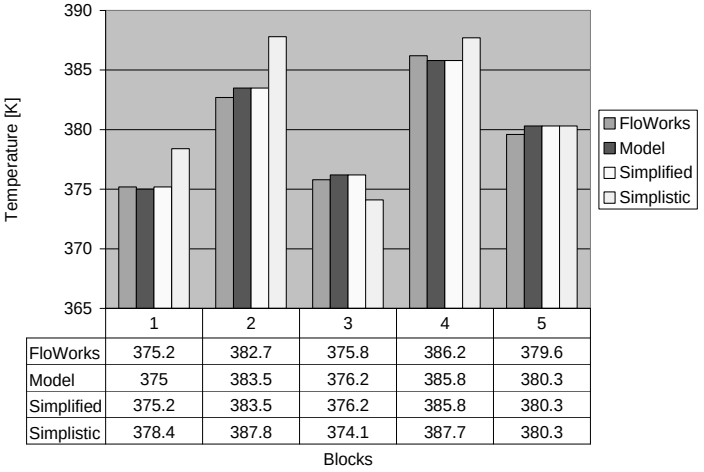


Figure 11. Layout #2



	1	2	3	4	5
FloWorks	375.2	382.7	375.8	386.2	379.6
Model	375	383.5	376.2	385.8	380.3
Simplified	375.2	383.5	376.2	385.8	380.3
Simplistic	378.4	387.8	374.1	387.7	380.3

Figure 12. Steady-state Average Temperature for Layout 1 and Power 1.



	1	2	3	4	5
FloWorks	375.2	382.7	375.8	386.2	379.6
Model	375	383.5	376.2	385.8	380.3
Simplified	375.2	383.5	376.2	385.8	380.3
Simplistic	378.4	387.8	374.1	387.7	380.3

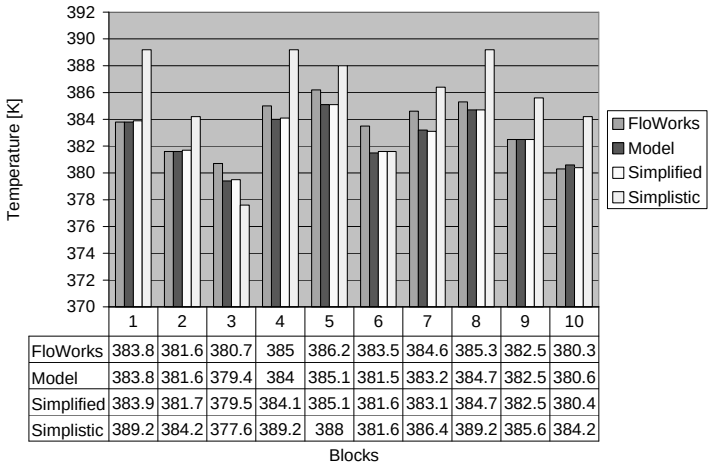
Figure 13. Steady-state Average Temperature for Layout 1 and Power 2.

Power for each block in Layout 2	Power1	Power1 density (watts/mm^2)	Power 2	Power 2 density (watts/mm^2)
Block1	30W	1.071	10W	0.357
Block2	20W	1.667	20W	1.667
Block3	10W	0.417	10W	0.417
Block4	15W	0.625	40W	1.667
Block5	25W	1.25	20W	1.667

Table 2. Power Densities in Layout 1.

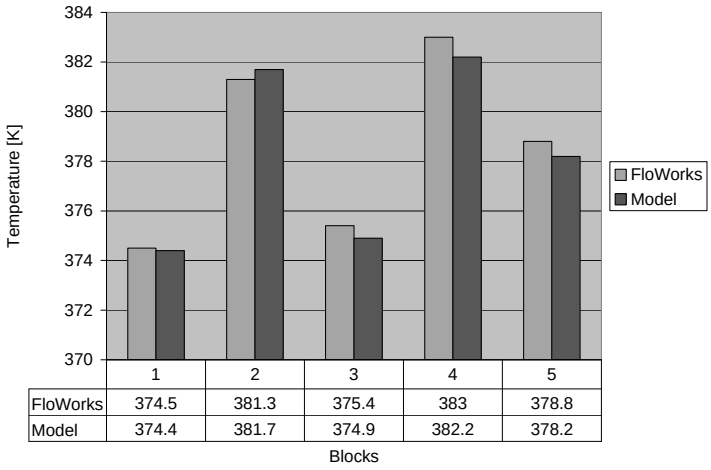
Power for each block in Layout 3	Power1	Power 1 density (W/mm^2)
Block1	10W	1.25
Block2	8W	0.8
Block3	2W	0.2
Block4	15W	1.25
Block5	15W	1.875
Block6	5W	1.25
Block7	12W	1
Block8	15W	1.25
Block9	13W	0.93
Block10	5W	0.5

Table 3. Power Densities in Layout 2.



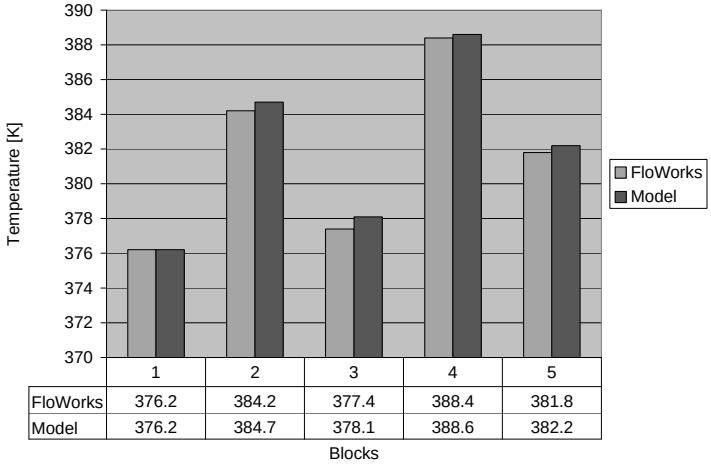
	1	2	3	4	5	6	7	8	9	10
FloWorks	383.8	381.6	380.7	385	386.2	383.5	384.6	385.3	382.5	380.3
Model	383.8	381.6	379.4	384	385.1	381.5	383.2	384.7	382.5	380.6
Simplified	383.9	381.7	379.5	384.1	385.1	381.6	383.1	384.7	382.5	380.4
Simplistic	389.2	384.2	377.6	389.2	388	381.6	386.4	389.2	385.6	384.2

Figure 14. Steady-state Average Temperature for Layout 2.



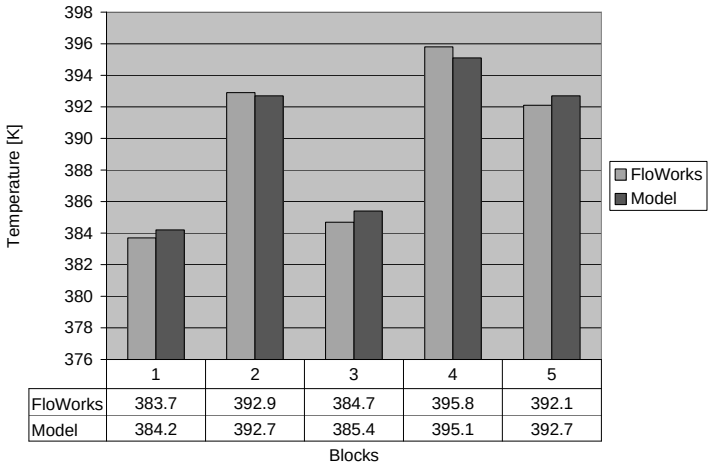
	1	2	3	4	5
FloWorks	374.5	381.3	375.4	383	378.8
Model	374.4	381.7	374.9	382.2	378.2

Figure 15. Steady-state Average Temperature for Layout 1 and Power 2 and Thickness = 0.3 mm.



	1	2	3	4	5
FloWorks	376.2	384.2	377.4	388.4	381.8
Model	376.2	384.7	378.1	388.6	382.2

Figure 16. Steady-state Average Temperature for Layout 1 and Power 2 and Thickness = 0.7 mm.

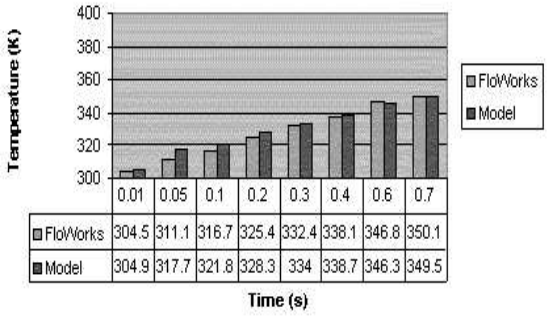


	1	2	3	4	5
FloWorks	383.7	392.9	384.7	395.8	392.1
Model	384.2	392.7	385.4	395.1	392.7

Figure 17. Steady-state Average Temperature comparison (2-2)

7. SIMULATIONS AND EVALUATION (TRANSIENT)

In FloWorks, we simulated temperatures at different time points for layout 1 with the power distribution 2. The model is represented in Figure 9, with the values of the resistors and capacitors gathered in Table 1. The short-term transient simulation results for Block 1 and Block 4 are shown in Figure 18 and Figure 19 respectively. A Cadence transient simulation of the compact RC model is shown in Figure 19. Typical thermal capacitances for heat sinks are around 300 J/K. Here, due to the very long simulation time, we assumed a capacitance of 1 J/K. The time dependence of temperature is shown in Figure 20, where the model is simulated in Cadence. This set-up is considered as a preliminary test of our model, further work needs to be done be done.



	0.01	0.05	0.1	0.2	0.3	0.4	0.6	0.7
FloWorks	304.5	311.1	316.7	325.4	332.4	338.1	346.8	350.1
Model	304.9	317.7	321.8	328.3	334	338.7	346.3	349.5

Figure 18. 1st Transient Average Temperature Comparison for Layout 1, Power2.

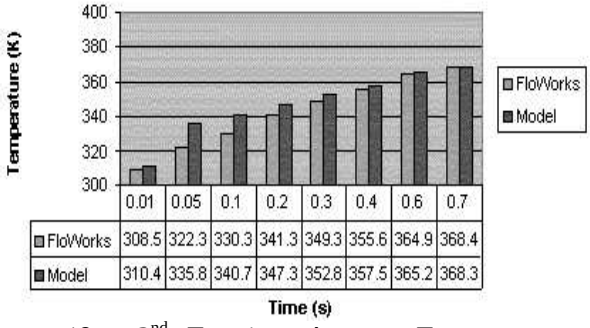


Figure 19. 2nd Transient Average Temperature Comparison for Layout 1, Power 2.

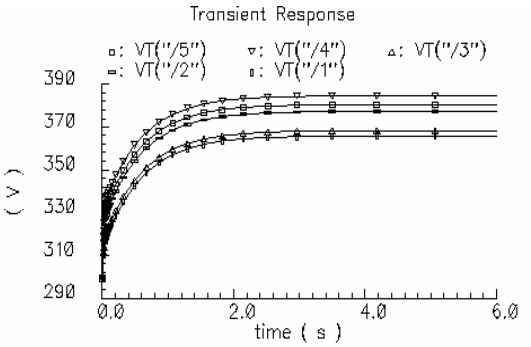


Figure 20. Model transient simulation for Layout 1 Power 2

8. CONCLUSION AND FURTHER WORK

Both the steady state and transient architecture level models are accurate enough (less than 4% error for the steady-state) and they are very compact and suitable for an electro-thermal simulator. The steady-state model is proposed in a simplified form as well. Both versions are very accurate and extremely small. They manage to capture the essential features of the heat flow, both in the vertical and lateral directions. A simplistic model, without lateral resistances showed that lateral heat flow is important and needs to be considered. The transient modeling needs further analysis, though. This includes considering a better way of modeling the lateral and vertical capacitors as well as testing the model thoroughly. The fact that such a small and fast model can be so accurate is an incentive to continue such an analysis. The models can be included into a cycle-based simulator, like Wattch (for power estimation) and SimpleScalar (for the performance estimation). The average temperature for each architectural block can be computed very accurately every cycle from our model, thus important information is available architecture level analysis such as layout cell placement. We used a numerical analysis tool, FloWorks, as the temperature reference for our model. FloWorks takes into account both the heat transfer in solid materials, like silicon and copper and the convection with the air, therefore, the results can be used as reference. Further work includes more simulations of transient thermal effects to verify our transient modeling techniques. The model can also be used for Dynamic Thermal Management (DTM) and architecture level analysis. [16] [17]

REFERENCES

[1] FloWorks. Dassault System, www.floworks.com

[2] H. Chiueh, J. Drapper, and J. Choma, Jr. "A Thermal Management System and Prototyping for System-on-chip Designs". *2001 Southwest Symposium on Mixed-Signal Design, SSMS.* pp 51 – 55, 2001.

[3] V. Koval and I. Farmaga. "MONSTR: A Complete Thermal Simulator of Electronic Systems". *31st ACM/IEEE Design Automation Conference.* pp. 570-575, 1994.

[4] V. Szekely and M. Rencz. "Fast Field Solver Programs for Thermal and Electrostatic Analysis of Microsystem Elements". *Computer-Aided Design, 1997. Digest of Technical Papers, 1997 IEEE/ACM International Conference.* pp. 684-689, 1997.

[5] H. Chiueh, J. Drapper, L. Luh, and J. Choma, Jr. "A Novel Model for On-chip Heat Dissipation". *Circuits and Systems, 1998. IEEE APCCAS 1998. The 1998 IEEE Asia-Pacific Conference*, pp 779-782, 1998.

[6] M. Janicki and A. Napieralski. "Temperature Monitoring of Electronic Circuits Based on Analytical Thermal Model". *Microelectronics, 2000. Proceedings. 2000 22nd International Conference, Vo2,1999.* pp 581 –584.

[7] W. Batty, C. E. Christoffersen, S. David, and A. J. Panks. "Fully Physical Time-Dependent Compact Thermal Modeling of Complex Non Linear 3-Dimensional Systems for Devices and Circuit Level Electro-Thermal CAD". *Semiconductor Thermal Measurement and Management, 2001. Seventeenth Annual IEEE Symposium, 2001.* pp 71 –84.

[8] C. Tsai and S. Kang. "Substrate Thermal Model Reduction for Efficient Transient Electro-thermal Simulation". *Mixed-Signal Design, 2000. SSMSD. 2000 Southwest Symposium, 2000.* pp 185 –190.

[9] J. T. Hsu and L. Vu-Quoc. "A Rational Formulation of Thermal Circuit Models for Electro-thermal Simulation—Part I: Finite Element Method". *Circuits and Systems I: Fundamental Theory and*

Applications, IEEE Transactions, Vol. 43 Issue 9, Sept. 1996, pp 721 –732.
[10] J. T. Hsu and L. Vu-Quoc. "A Rational Formulation of Thermal Circuit Models for Electro-thermal Simulation—Part II: Model Reduction Techniques". *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions, Vol 43 Issue 9, Sept. 1996*, pp 733 -744.
[11] C. Tsai and S. Kang. "Cell-Level Placement for Improving Substrate Thermal Distribution". *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions, Vol 19 Issue 2 , Feb. 2000*. pp 253 –266.
[12] M. Rencz and V. Szekely. "A Generic Method for Thermal Multiport Model Generation of IC Packages". *Semiconductor Thermal Measurement and Management, 2001. Seventeenth Annual IEEE Symposium, 2001*. pp 145 –152.
[13] A. Aranyosi, A. Ortega, and J. Evans. "Development of Compact Thermal Models for Advanced Electronic Packaging: Methodology and Experimental Validation for A Single-Chip CPGA Package". *Thermal and Thermomechanical Phenomena in Electronic Systems, 2000. ITHERM 2000. The Seventh Intersociety Conference, 2000.* pp 225-232.
[14] G. Noebauer. "Creating Compact Models Using Standard Spreadsheet Software". Semiconductor Thermal Measurement and Management, 2001. Seventeenth Annual IEEE Symposium, 2001. pp 126 –133.
[15] Intel microprocessor thermal specifications and guidelines: http:// www.stwing.upenn.edu/~rkuo/heatspec.html
[16] Kevin Skadron et al.. "Control-Theoretic Techniques and Thermal-RC Modeling for Accurate and Localized Dynamic Thermal Management" *Proceedings of the Eighth International Symposium on High-Performance Computer Architecture*, Feb. 2-6, 2002.
[17] David Brooks and Margaret Martonosi. "Dynamic Thermal Management for High-Performance Microprocessors" *Proceedings of the Seventh International Symposium on High-Performance Computer Architecture*, January 2001.