Fast Thermal Simulators for Architecture Level Integrated Circuit Design

Amirkoushyar Ziabari, Ehsan K. Ardestani, Jose Renau, and Ali Shakouri*, *member IEEE*

***Abstract*—High temperatures and non-uniform temperature distributions have become a serious concern since they limit both performance and reliability of Integrated Circuits (IC). With computer architect's concern to position microarchitecture blocks in a processor, faster thermal models can be developed at the cost of hiding finer grain details such as circuit or transistor level information. Several methods to quickly estimate the surface temperature profiles of microarchitecture blocks have been investigated in recent years. HotSpot simulator is widely used in computer architecture community. SESCTherm is another architecture level thermal simulator which has shown good performance and modularity in modeling. Recently Power Blurring (PB) method has been developed for both steady-state and transient thermal analysis of standard and 3D chips. While some of these methods are validated against finite element and Green's function based techniques, there are no detailed comparisons of the accuracy and speed for some common applications. In this paper we present the steady-state and transient temperature distributions calculated by these three architecture level thermal simulators. A detailed comparison taking into account the accuracy and the computation speed is performed. Our results indicate that Power Blurring has the potential to be a promising architecture level thermal simulator for fast calculation of temperature profile from the input power map in a realistic package which, in turn, is a key ingredient for full self-consistent simulations.**

***Index Terms*—Power Blurring, SESCTherm, HotSpot, thermal simulation, architectural level thermal simulator.**

I. Introduction

Increasing on-chip power density due to the shrinking of transistor size in the CMOS VLSI has resulted in elevated temperature in integrated circuit (IC) chips. Temperature non-uniformity across the chip leads to hot spots. This effect and the elevation in temperature are critical issues because of their impact on both the performance and the reliability of IC chips [1].

In addition, the increasing leakage power and its exponential dependence on the temperature require more attention to thermal-aware simulations and optimizations. Hence precise thermal profile is essential for accurate analysis of performance, reliability, and power management.

Generally, thermal simulations and design optimizations have been done under steady-state worst case conditions due to the high computational cost, causing reliance on the use of conservative margins in thermal designs. However, temperature non-uniformity evolves over time and hot spots can be transient. As the thermal budget becomes increasingly tight, the *worst case* approach becomes too costly and ineffective. Even with the state-of-the-art tools, chip-level transient thermal simulation with realistic package configuration is too expensive for physical design optimization or performance verification in the packaged environment. Additionally, in early stage of chip design, specific package information and thermal boundary conditions may not be available. In such cases, a fast thermal analysis method is highly desired [2]. Fast thermal simulations also allow optimization of the chip during architecture evaluation.

Our experiments show that in a fast power and thermal simulation for a processor, thermal simulation takes up to 20% of the time budget for a single core system. A multicore system might need even more time for thermal simulation. Hence accuracy as well as speed of the thermal model directly impacts the productivity in the evaluation phase of processor design.

This work provides the base for comparison of accuracy and speed of three different thermal simulation methods. We study three fast thermal analysis methods, namely, HotSpot [3], SESCTherm [4], and Power Blurring [5]. We provide a detailed comparison between the methods in both steady-state and transient thermal analysis.

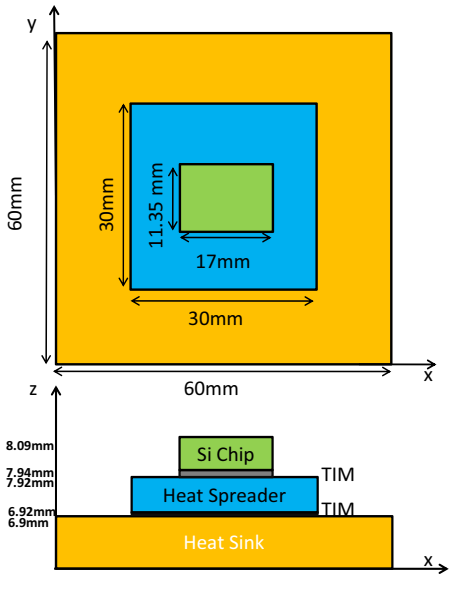
The reminder of the paper is organized as follows. Section II, provides a background for PB and other simulators. In section III, a steady-state and a transient case study is discussed and the results of the three methods are compared. We conclude the paper in section IV.

II. Background

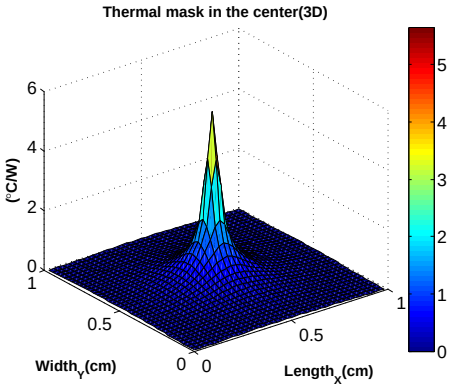
A. Power Blurring method

The PB method has its theoretical basis on the Green's function method and the methodological basis on image blurring used in image processing. Given a power distribution map of an IC, it can be thought of as an image, and the temperature distribution of the IC chip can be regarded as a blurred image of the power map. To perform the image blurring, a filter mask is required. A filter mask is equivalent to an impulse response, which is nothing but the Green's function of the system. The impulse response can be obtained by using Finite Element Analysis (FEA) tool such as ANSYS. To create a mask, a point heat source is applied to the center of the chip. An example of a packaged IC chip structure and the thermal mask are shown in Fig.1. The thermal profile for a given power map is obtained when the thermal mask is convolved with the given power map. The advantage of the PB technique is that it can be applied to realistic chip geometries where analytical Green's function cannot be obtained.

For a complicated power dissipation profile, the PB method reduces the calculation time by three orders of magnitude



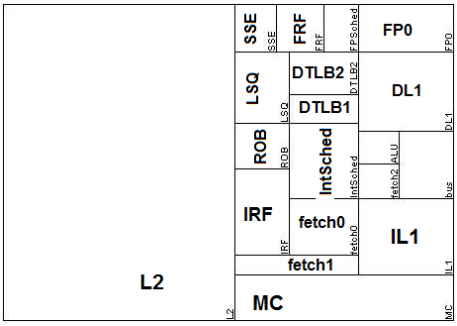
(a)



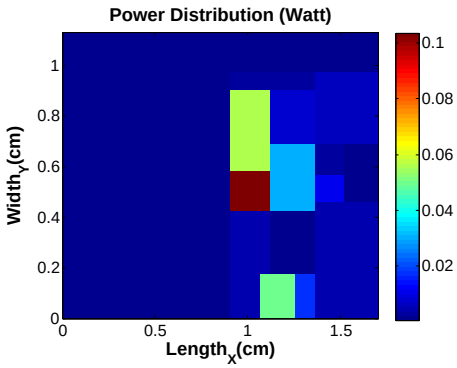
(b)

Fig. 1: (a) Packaged IC chip where the heat spreader, the heat sink and the thermal interface layers are included. (b) The thermal mask when a heat source is applied at the center of the chip.

compared to the Finite Element Analysis (FEA) [6], [7]. Since a point heat source at the center and that at the edge of the IC surface produce different peak temperatures and slightly different temperature profiles, usually three FEA simulations are needed in order to implement the power blurring technique for an arbitrary heat dissipation profile. The PB method was further improved by applying the Method of Image which takes into account the symmetry of the heat dissipation in a finite size chip [6]. Additionally, another error reduction step, deemed Intrinsic Error Compensation step, is used to improve the results [6]. This takes into account the three-dimensional geometry of the heat sink and the fact that a uniform power dissipation profile does not produce a constant temperature distribution everywhere in the chip. The PB method has been used in both steady-state and transient thermal simulations and is validated against experimental methods [6], [8], [9]. To consider the temperature dependency of silicon thermal



(a)



(b)

Fig. 2: The floorplan and power map for steady-state case study. (a) the floorplan of the device. (b) a typical mobile processor model power map.

conductivity, the method is extended to a self-consistent procedure or Adaptive Power Blurring [10]. It has also been shown that the temperature profile with resolution comparable to a single transistor stage/logic block size ($5\times5\ \mu m^2$) and for one million mesh points on the surface of chip can be acquired [7]. At such a fine resolution, FEM method such as ANSYS would be often limited by the number of mesh points and the computation times will be very long.

TABLE I: Material Properties and Dimensions of the package model.

	Area (mm^2)	Thickness (mm)	Thermal Conductivity (W/m-K)	Density (kg/m^2)	Specific Heat (J/kg-K)
Si Die	17×11.35	0.15	100	2330	751
TIM1	17×11.35	0.020	4	1930	2072.5
Heat Spreader	30×30	1	400	8933	397.4
TIM2	30×30	0.020	4	1930	2072.5
Heat Sink	60×60	6.9	400	8933	397.4

B. *HotSpot*

Architecture level simulators are designed to calculate temperature profiles which are accurate for the experiments at architecture level (block sizes in millimeter range), and still fast enough to allow for simulation of long dynamic

TABLE II: Comparison between the three methods.

	SESCTherm	Hotspot	PB
Computation Time	-	0.11s	0.041s
Err. in hot-spot	13.1%	12.9%	0.14%
Max. Err.	43.7%	25.7%	13.7%
Avg. Err.	15%	6.5%	2.5%
Abs. Err. range	0-5.3°C	0-4.2°C	0-0.56°C

temperature traces on the order of seconds. Their main feature, small computation time compared to detailed finite element models, comes at the cost of accuracy. However, this allows architects to study thermal and performance trade-offs in their system design. For that, a lot of details typically considered in a full thermal design, is deliberately neglected.

HotSpot is one of the thermal simulators widely used in computer architecture community. It is based on an equivalent circuit of thermal resistances and capacitances that correspond to the micro-architecture blocks. The essential aspects of the thermal package are also taken into account [3]. HotSpot solves the heat differential equations describing the RC circuit at each time step using a fourth-order Runge-Kutta method. The number of iterations for the RungeKutta solver is adaptive, to account for the different number of iterations required for convergence at different sampling intervals. HotSpot is already configurable for purposes of modeling new floorplans. HotSpot can model steady-state as well as transient cases. It can be run in two levels of accuracy: block model and grid model. While block level simulation has higher speed, the grid model is more accurate due to smaller spatial granularity of elements. We compare our method with the HotSpot in grid mode.

C. SESCTherm

SESCTherm is a thermal modeling infrastructure based on finite-difference analysis techniques. At the core of SESCTherm is a finite-difference model (FEM). Finite-difference analysis involves taking a problem and segmenting the problem into smaller pieces. SESCTherm divides the chip, package and associated components into a series of regular cross sectional regions. Each region is a quadrilateral, and no two cross sectional regions have abutting sides that are of different height or length. Each cross-sectional region is called a temperature node, and each region has a series of properties. To accurately characterize complex materials and dimensions, SESCTherm subdivides regions by material and geometry. This means that each temperature node is considered to be one material or an approximation of a combination of materials. Further, this means that any irregular shape is approximated by a combination of quadrilateral regions. Each node of a thermal system can be considered either a heat source, heat sink, thermal capacitance, or thermal resistance. SESCTherm also updates the material properties as the temperature changes. It could support stacked layers of die and has models for interconnect layer, package and mainboard, as well as bulk silicon and silicon-on-insulator. To model temperature dependent material properties, SESCTherm updates material properties periodically based upon temperature variations. It supports configurable spatial granularity for different levels of accuracy. Unlike HotSpot, The meshing system is a consolidation of grid and block modes in which floorplan edges are extended and then each resulted region is meshed based on a given granularity [4]. However, the cores of the differential equation solver for both HotSpot and SESCTherm are based on the same algorithm.

III. Case Studies

To compare the three methods, we used two floorplans. One models the power dissipation profile of a typical out-of-order mobile processor, and the other one is a simple 4×4 grid. We configured each tool with the same parameters in terms of heat-sink and package characteristics.

A. Error Metrics

In order to study the accuracy of each method, we calculated the relative error compared to that of ANSYS, which is a standard FEM tool for thermal analysis, using Equation. 1. In all simulations ambient temperature is set to 35°C.

$$Err = (T_{Method} - T_{ANSYS}) \,/\, (T_{ANSYS} - T_{Ambient}) \quad (1)$$

Max. Error: For each grid across the entire chip Equation. 1 is used to calculate the error and then the maximum error is reported.
Hot-spot Error: Equation. 1 is used to calculate the error in the hottest spot in the temperature profile.
Average Error: Equation. 1 is applied to the average temperature across the chip.
Absolute Temperature Error range: This error is obtained using Equation. 2 and then the range is reported.

$$Err = (T_{Method} - T_{ANSYS}) \quad (2)$$

B. Methodology Validation

The convection coefficient between the heat sink and air is obtained using Equation. 3. In this equation R is the convection resistance between the heat sink and air, A is the surface area of the heat sink, and h is the equivalent convection coefficient.

$$h = 1/(R \times A) \quad (3)$$

The convection resistance between heat sink and air is 0.1 K/W, and the surface area of the heat sink is 36 cm^2. This results in a convection coefficient of 0.2778 W/m^2K. For ANSYS and Power Blurring we used this value of convection coefficient for our simulations. To make sure the overall chip and package models for all four methods match, and make the comparison fair, we perform one step of calibration before doing the evaluation. For that, instead of setting the parameters in HotSpot and SESCTherm to match the 0.2778 W/m^2K, we try to adjust the convection coefficient to such a value that the overall average error is minimized. Then we evaluated the relative error values. For example in the steady-state

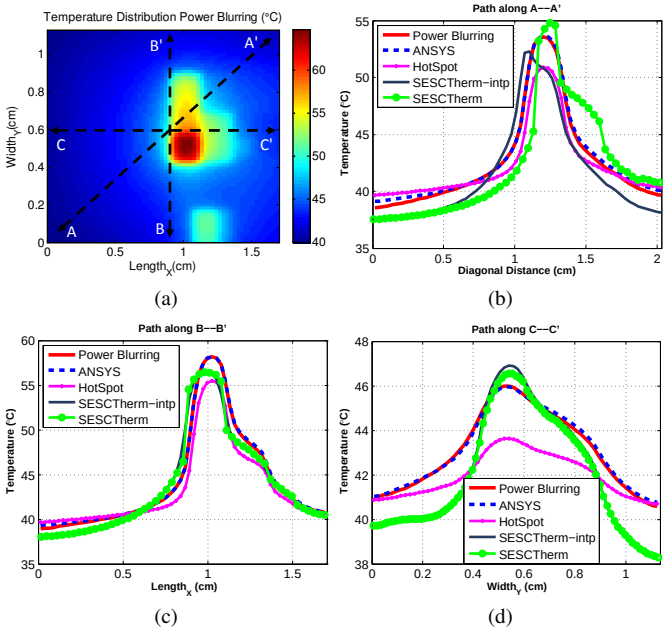


Fig. 3: A steady-state comparison between ANSYS, PB, HotSpot, and SESCTherm (both row and interpolated temperature profiles). (a) The temperature profile obtained by the PB method corresponding to the power map shown in Fig. 2b. (b) Temperature profiles over the diagonal of the chip. (c) Temperature profiles along the device length. (d) Temperature profiles along the device width.

case study, the initial average error obtained for HotSpot and SESCTherm were 19 % and 26 %, respectively. After applying this adjustment procedure, the average errors are minimized to 6 % and 15 %, which is shown in next section(Table.II). Then for the same optimized parameters we obtained the transient case study results.

In order to achieve the minimum average temperature error in HotSpot, we explicitly changed the value of convection resistance to 0.13 K/W. This is equivalent to 0.2137 W/m^2K for convection coefficient which is a factor of 1.3 smaller than its real value (0.2778 W/m^2K). In the case of SESCTherm, the error reduction procedure was different. SESCTherm does not allow the user to explicitly specify a value for convection resistance. Instead, it computes this value based on the given geometric and material properties, as well as the parameters of the cooling solution.

C. A steady-state case study

For steady-state thermal comparison, we used the power map from a typical mobile processor model. The dimensions of the chip and its cooling solution are set according to Fig. 1 and Table I. The power map, the floorplan of the device as well as the results obtained by three methods are shown in Fig. 2 and Fig. 3, respectively. It should be mentioned that in order to obtain these results, a 64×64 meshing size was used, which is a default grid size in HotSpot [11]. This meshing also results in 266×177μm^2 granularity which is fairly fine-grained for thermal evaluation of a processor at architecture level.

Table II summarizes the computation time, the absolute temperature error range, maximum error, the average error, and error in the hottest spot of the chip for each of the PB, HotSpot and SESCTherm methods. From Fig. 3 and Table II, it can be concluded that the PB method offers a more accurate result while its execution time is shorter than the SESCTherm and HotSpot. The latter delivers a more accurate result compare to SESCTherm, although it cannot provide the temperature profile with accuracy and resolution of the PB method. Since SESCTherm does not support a separate method for steady state analysis, we do not report the execution time for this case. In order to obtain the static result using SESCTherm, we have applied the power input and obtained the temperature profile after a long period of time to be able to consider it as steady-state response. The computation time for ANSYS is 56s. In ANSYS we have used the sparse equation solver algorithm in which the time complexity is of the order of O (n^2), for a thermal circuit with n nodes, while the time complexity of the FFT algorithm used in PB method is of the order of O (nlog(n)). SESCTherm and HotSpot use traditional integration based solvers for which the lower bound of the computation time complexity is of the order of O (n^c) where c is a number between 1.5 to 2 [12]. Considering these orders, one can see that by increasing the number of nodes in thermal grid model, the computation time of the PB method increases with asymptotically slower rate than the other methods.

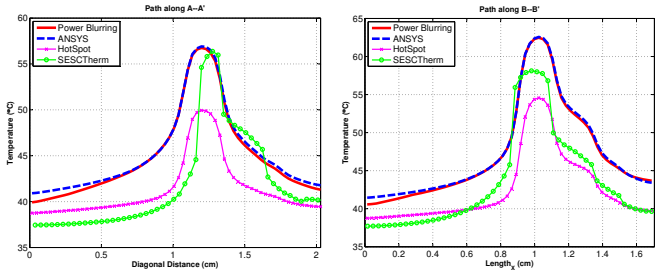


Fig. 4: Steady state comparison between four methods using nominal value of convection coefficient (0.2778 W/m^2K). (a) Temperature profiles over the diagonal of the chip. (b) Temperature profiles along the device length

The error of PB compare to ANSYS in the hottest spot of the chip is only 0.14%, while it is 12.9% for HotSpot, and 13.1% for the SESCTherm simulator. The maximum error of 13% for the PB method relative to ANSYS in the entire profile is due to a temperature difference of 0.6°C (39.1-38.5). Since this small change of temperature occurs at the very edge of the chip, in which the temperature is much lower than the center and very close to ambient temperature (35°C), it will result in a large error value (see Equation. 1) even though it is in fact a negligible change. The maximum errors in the HotSpot and SESCTherm simulators are 4 and 5 degrees temperature differences, respectively. The average error and absolute temperature error range in Table II indicate that the PB method accurately estimates the temperature distribution throughout the chip.

All the aforementioned error values are obtained after adjustment of convection coefficient in HotSpot and SESC-

Therm. In Fig. 4a and Fig. 4b a comparison between the cross sections of steady state results of these methods, with the nominal value of convection coefficient i.e. 0.2778 W/m^2K, is presented.

As mentioned, the procedure of meshing in SESCTherm is different from other methods. Based on a known granularity, structure of the blocks, and aspect ratio of the chip, SESCTherm automatically determines the meshing. In this case study, the meshing size, for the specified 230μm spatial granularity, is 79×53. This results in an asymmetric plots for the diagonal view of the temperature profile as it can be seen in Fig. 3b. To be able to calculate the relative error, the matrix dimensions of temperature profiles must be the same. Therefore, we interpolated the temperature profiles obtained by SESCTherm to 64×64 matrices. This, in turn, might add some inaccuracy to the relative error results obtained by SESCTherm, even though it does not change the average error. The cross sections of the interpolated result are also indicated in Fig. 3.

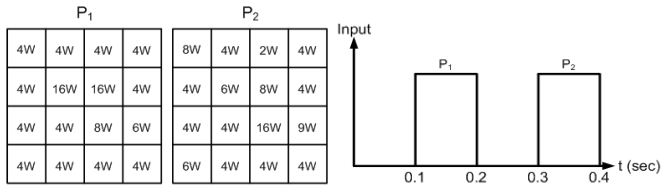


Fig. 5: Pulse input of the coarse power maps: (a) coarse power maps. (b) power dissipation pattern.

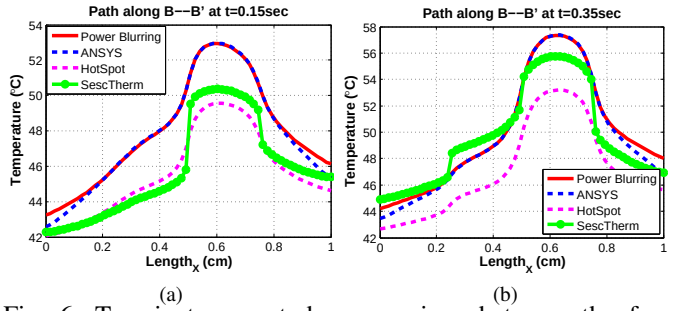


Fig. 6: Transient case study comparison between the four methods. Thermal profile B-B at: (a) t=0.15s; and (b) t=0.35s.

D. A transient case study

Transient simulations are performed for a power train input. The coarse power maps are shown in Fig. 5. We used a 1×1 cm^2 chip with a cooling solution similar to the one shown in Fig. 1. The chip has been meshed with 64×64 grid size. The resulting temperature profiles at t=0.15s and 0.35s are presented in Fig. 6a, and Fig. 6b, respectively. In Table III a detailed comparison between the three methods is presented. The errors calculated in the table are for the time t=0.15s and 0.35s. The maximum error of 18% in PB corresponds to 1 °C absolute error. This occurs at the edge of the device where the temperature value is close to the ambient temperature and leads to a very small value in the denominator of the error function. A same argument is valid for large errors in HotSpot and SESCTherm and their absolute error values has to be considered which are provided in Table.III. The execution time for the PB method was 194 seconds while this value was 290, 697, and 27858 seconds for the SESCTherm, HotSpot, and ANSYS, respectively. As it can be seen the PB method is about 1.5, and 4 times faster than SESCTherm and HotSpot, respectively. PB also provides more accurate results. It should be mentioned that the PB method relies on two FEA simulations (or measurements) giving the unit impulse response at the center of the chip and the additional correction factor from uniform power dissipation profile. These calculations could be done offline, so the main advantage of PB is in multiple thermal simulations when different placements of the IC blocks are studied. All the matrix arithmetic calculations in that PB method have been done in MATLAB. While this is flexible and it allows the use of image processing tools, it is anticipated that direct implementation of the matrix convolution in a higher level program (e.g. C) can increase significantly the speed of the PB method.

TABLE III: Comparison between the three methods (Transient).

	SESCTherm	Hotspot	PB
Computation Time	290s	697s	194s
Err. in hot-spot @0.15s	3.9%	16.7%	0.24%
Max. Err. @0.15s	33.2%	22.2%	16%
Avg. Err. @0.15s	9.8%	11.7%	2.1%
Abs. Err. range @0.15s (°C)	0-6.7	0-4.7	0-1
Err. in hot-spot @0.35s	7.5%	16%	0.13%
Max. Err. @0.35s	39%	22.3%	18.6%
Avg. Err. @0.35s	7.8%	15.3%	3%
Abs. Err. range @0.35s (°C)	0-6.9	0-4.5	0-1.34

Finally, we have investigated the evolution of error over time in the PB method. As it can be seen in the Fig. 7, when the device is on and power being dissipated, the maximum absolute temperature error over the entire chip is less than 2 °C. Furthermore, it is shown that the absolute error in the hottest spot over the entire chip with power blurring method is less than 0.2 °C.

IV. Conclusion

Fast and accurate thermal model enables in depth thermal evaluation in processor design. This work compares three different thermal models, namely Power Blurring (PB), HotSpot and SESCTherm. HotSpot and SESCTherm are two standard architecture level simulators, and PB method has been recently proposed for thermal simulations. To perform the comparison, we adjusted the convection coefficient to such a value that the overall average errors in HotSpot and SESCTherm are minimized. This adjustment is a fitting parameter which does not have a scientific justification and could be different for different packages or even for different power profiles. Having validated the comparison methodology, both steady-state and transient case studies have indicated that the PB method can provide more accurate temperature profiles with

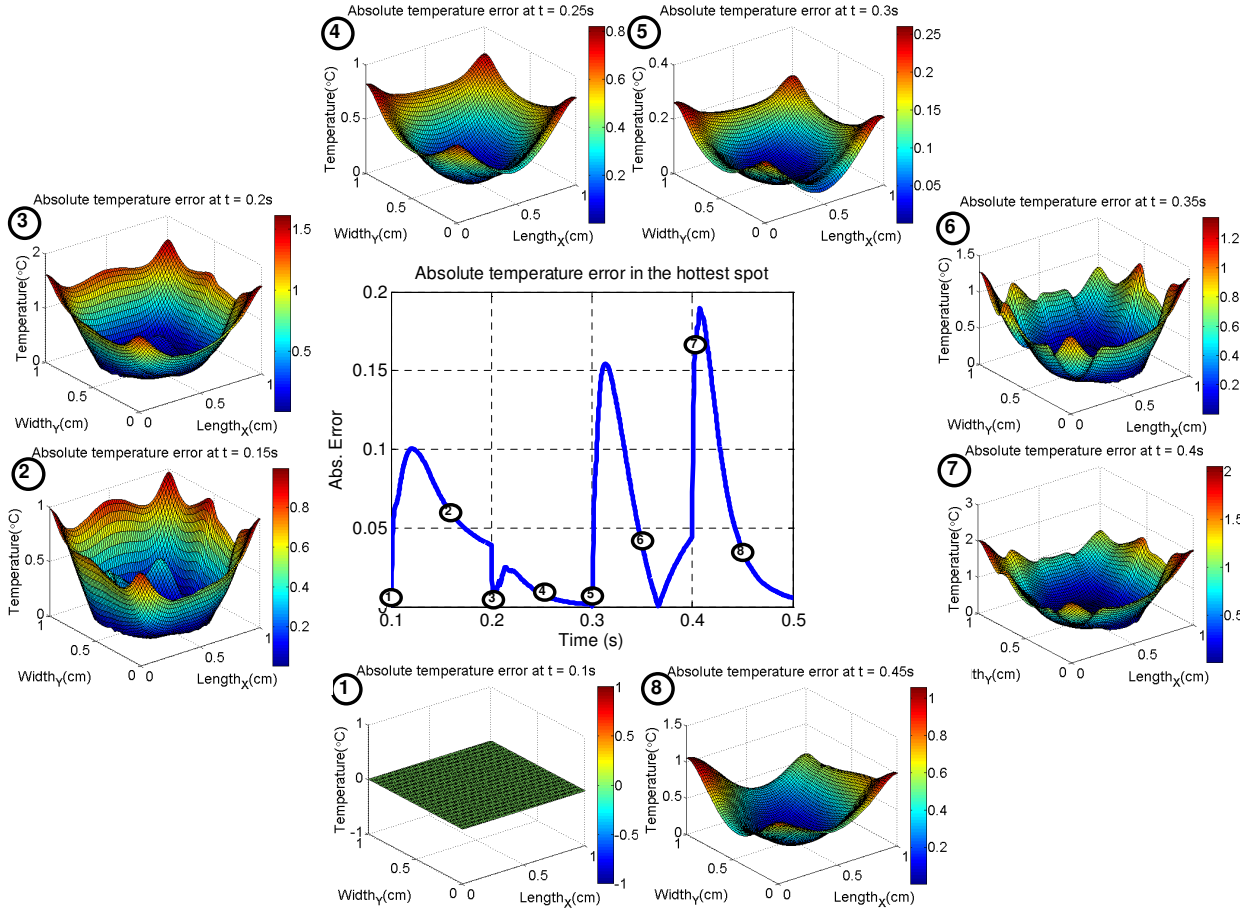


Fig. 7: Center image shows the evolution of absolute temperature error in the hottest spot over the entire chip. (1-8) Absolute temperature error profiles at different times over the chip.

shorter execution times. This is advantageous for in depth exploration of trade-offs in early stages of processor design process. Another application is in very high precision thermal simulation with micron scale power dissipation profile. Computation time for power blurring which is a surface(2D) matrix convolution scales much better than the volume 3D meshing techniques. These results demonstrate that power blurring has the potential of becoming a promising thermal simulation tool in architecture level.

V. Acknowledgment

We would like to thank Dr. Mircea Stan and Dr. Kevin Skadron for their help and feedbacks on this work. This work was supported in part by the Interconnect Focus Center, one of the research centers funded under the Focus Center Research Program, a DARPA and Semiconductor Research Corporation program.

References

[1] A. H. Ajami, K. Banerjee and M. Pedram, *Modeling and Analysis of Non-uniform Substrate Temperature Effects on Global ULSI Interconnects*, IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 24, no. 6, pp. 849-861, 2005.

[2] Y. K. Cheng and S. M. Kang, *A Temperature-Aware Simulation Environment for Reliable ULSI Chip Design*, IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 19, no. 10, pp. 1211-1220, 2000.

[3] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, *Temperature-Aware Computer Systems: Opportunities and Challenges*, IEEE Micro, vol. 23, no. 6, pp. 52-61, 2003.

[4] J. Nayfach and J. Renau, *SOI, Interconnect, Package, and Mainboard Thermal Characterization*, International Symposium on Low Electronics and Design (ISLPED), pp. 327-330, Aug. 2009.

[5] T. Kemper, Y. Zhang, Z. Bian, and A. Shakouri, *Ultrafast Temperature Profile Calculation in IC chips*, Proc. of 12th International Workshop on Thermal investigations of ICs (THERMINIC), Nice, France, pp. 133-137, 2006.

[6] V. M. Heriz, J. H. Park, A. Shakouri, and S. -M. Kang, *Method of Images for the Fast Calculation of Temperature Distributions in Packaged VLSI Chips*, Proc. of 13th International Workshop on Thermal investigations of ICs (THERMINIC), Budapest, Hungary, September 17-19, 2007.

[7] J. H. Park, X. Wang, A. Shakouri, and S. -M. Kang, *Fast Computation of Temperature Profiles of VLSI ICs with High Spatial Resolution*, Proc. 24th Semi-Therm, San Jose, CA, pp. 50-54, 2008.

[8] J. H. Park, A. Shakouri, and S. -M. Kang, *Fast Evaluation Method for Transient Hot Spots in VLSI ICs in Packages*, International Symposium on Quality Electronic Design, San Jose, CA, pp. 600-603, 2008.

[9] J. H. Park, S. Shin, J. Christofferson, A. Shakouri, and S. -M. Kang, *Experimental Validation of the Power Blurring Method*, Proc. 26th Semi-Therm, Santa Clara, CA, pp. 240-244, February 21-25, 2010.

[10] A. Ziabari, Z. Bian, and A. Shakouri, *Adaptive Power Blurring Techniques to Calculate IC Temperature Profile under Large Temperature Variations*, International Microelectronic and Packaging Society (IMAPS) ATW on Thermal Management, Sept 28-30, 2010.

[11] http://lava.cs.virginia.edu/HotSpot/index.htm

[12] P. Liu, Z. Qi, H. Li, L. Jin, W. Wu, S. X. -D. Tan, and J. Yang, *Fast Thermal Simulation for Architecture Level Dynamic Thermal Management*, IEEE/ACM International Conf. on Computer-Aided Design, pp. 639-644, 2005.