

Two Fast Approaches for 3D Thermal Simulation of Integrated Circuits

[Invited Special Session Paper]¹

Wenjian Yu

Dept. Computer Science and Technology, Tsinghua University, Beijing 100084, China

Email: yu-wj@tsinghua.edu.cn

Abstract

To perform accurate chip-level thermal simulation, the irregular 3D structure including heat sink components is considered. Two approaches, i.e. domain decomposition and hybrid random walk, are presented to accelerate the tasks of calculating entire temperature profile and hot-spots temperatures, respectively. They both are much faster than existing techniques while keeping good accuracy, and suitable for the high-resolution thermal simulation for design optimization and verification.

1. Introduction

The continuous scaling trend of the CMOS technology has led to the drastic increase of the number of devices in integrated circuit (IC). The heat dissipation has become a problem that threatens circuit reliability and performance [1]. The chip-level thermal analysis is performed during the sign-off stage for performance and reliability verification. Accurate and efficient thermal analysis is also indispensable for many design-time circuit optimizations, where a large number of thermal simulations with different power dissipation distributions are required. Now, three-dimensional (3D) IC has been developed to reduce the interconnect delay and enable the heterogeneous integration [2]. The increased power density and the low-thermal-conductivity inter-layer material in 3D IC cause a more serious problem of heat dissipation. This further increases the importance of accurate thermal simulation during the IC design stage.

To address the problem of accurate thermal simulation, the whole thermal system, including the die of IC, the copper heat spreader attached to the die and the heat sink attached to the spreader, should be considered (see Fig. 1). Note that the simulation only involving the die will incur significant error of temperature [3].

Some algorithms have been proposed for chip-level thermal analysis, such as the geometric multigrid solver [4], Green's function based fast algorithm [5], and the preconditioned conjugate gradient (PCG) algorithm [3]. The volume discretization has been employed to transform

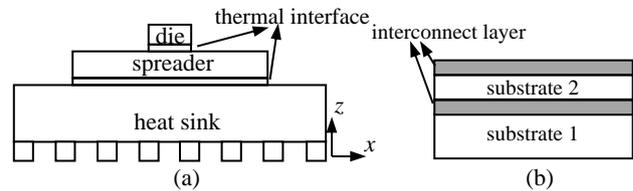


Fig. 1. The pyramid-shape IC geometry for thermal simulation: (a) the side view, (b) the details of a IC with two tiers of dies.

the problem into a linear equation system, and the temperature profile of the whole simulated domain is solved. Most of these works only consider a rectangular simulation domain, with a simplified boundary assumption (often the Dirichlet condition) accounting for the heat dissipation effect. When the real pyramid-shape model is considered, a large number of unknowns involved will cause excessive runtime and memory cost.

Sometimes, the entire temperature profile is not required. What we really concern is the temperature of hot-spots at IC device layer. For this purpose, the random walk method is suitable. The random walk algorithms have been proposed for power grid simulation [6, 7]. Based on the similarity of thermal analysis and power grid simulation, the random walk method [6] was applied to the problems of thermal via planning [8] and thermal analysis [9]. The random walk method has also been successfully applied to problem of capacitance extraction [10, 11].

In this paper, two approaches are presented to accelerate the chip-level thermal simulation. Firstly, a technique based on the domain decomposition method (DDM) [12] is briefly introduced. It alleviate the problem of simulating entire temperature profile while considering the pyramid-shape IC model. Then, a hybrid random walk (HRW) algorithm is presented, which is suitable for calculating the temperature of some hot-spots. HRW is more general for handling complex IC thermal models. Both approaches are orders of magnitude faster than existing techniques, while preserving good accuracy.

2. Background

The steady-state thermal analysis involves solving the temperature $T(x,y,z)$ from the 3-D Poisson equation:

$$k \left[\frac{\partial^2 T(x,y,z)}{\partial x^2} + \frac{\partial^2 T(x,y,z)}{\partial y^2} + \frac{\partial^2 T(x,y,z)}{\partial z^2} \right] = -p(x,y,z), \quad (1)$$

where k is thermal conductivity and $p(x,y,z)$ is the internal heat generation density at point (x,y,z) . The heat generation is due to the device modules, or function blocks

¹ This paper is invited by Special Session of "Advanced CAD Techniques for Power/Temperature-Aware IC Design". The work was supported in part by National Natural Science Foundation of China (No. 61076034), the Opening Foundation of ASIC and System State Key Laboratory (No. 12KF009), and the Tsinghua University Initiative Scientific Research Program.

located around the top surface of the silicon die. Eq. (1) holds for a homogeneous region. For a problem with multiple homogeneous regions, the equation of continuous heat flux should be applied at the region interfaces.

The finite volume method (FVM) is conventionally used for 3-D thermal simulation, where the domain is discretized into cells and each cell is associated with a temperature [1, 3, 4, 8, 9, 12]. Similar to simulating the steady-state electric current field with electric resistors, we can define and calculate thermal resistor to model the heat flow through the interface between any two adjacent cells. Fig. 2 illustrates the finite volume discretization of a 3-D rectangular domain. For two cells with different thermal conductivities or different lengths along the aligned direction, the thermal resistance connecting them can be calculated as the series of resistors. For example, the R_{tz} shown in Fig. 2(c) can be calculated with:

$$R_{tz} = \frac{1}{2k_1} \cdot \frac{h_{z1}}{h_x h_y} + \frac{1}{2k_2} \cdot \frac{h_{z2}}{h_x h_y}, \quad (2)$$

where h_x and h_y are edge sizes of cell along x -axis and y -axis, respectively. h_{z1} and h_{z2} are the heights of the two adjacent cells; k_1 and k_2 are the thermal conductivities.

The heat source resembles the current source in electric circuit. Thus, an equivalent resistor circuit is generated, which corresponds to a linear equation system:

$$AT = f, \quad (3)$$

where A is a sparse symmetric positive definite matrix, f is the vector of current sources, and T is the temperature vector. The temperature can be obtained by solving (3).

There are different boundary conditions for the simulation domain. At the bottom surface of heat sink, a convective condition should be set, which models the heat transfer mechanism at the interface of heat sink and air:

$$k \frac{\partial T}{\partial \bar{n}} + h(T - T_{amb}) = 0, \quad (4)$$

where \bar{n} is the out normal direction of the boundary, T_{amb} is the ambient temperature, and h is the convective coefficient. The partial derivative in (4) can be approximated with finite difference formula. By defining $R_{amb} = 1/(h \cdot h_x \cdot h_y)$,

where h_x and h_y are the edge sizes of cell along x -axis and

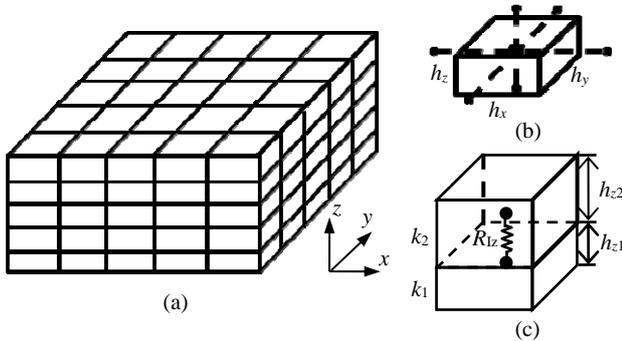


Fig. 2. The illustration of FVM and thermal resistors. (a) The discretization of a homogeneous material. (b) A cell and its six thermal resistors. (c) A thermal resistor across material interface.

y -axis respectively, the effect of convective boundary is modeled with thermal resistors of R_{amb} . They connect the nodes of boundary cells to a virtual node with temperature T_{amb} . For other boundaries of the domain, Neumann boundary (adiabatic condition) is usually assumed. It can be modeled by the equivalent circuit naturally.

3. The Domain-Decomposition Approach

DDM is a general method for simulating problems with complex geometry or topology, and has been applied to circuit simulation problems [13]. We find out that it has distinct merit for the thermal simulation considering the whole pyramid-shape model. For the thermal model shown in Fig. 1, it is nature to divide the whole domain into three subdomains representing the chip, heat spreader and heat sink respectively. With DDM, the subdomains are simulated separately, and therefore two advantages are brought. 1). The subdomain often has geometry regularity, which enables fast simulating algorithms [3]; 2). Because the power distribution and temperature in chip is only concerned, simulating the subdomains with different discretization resolutions can speed up the computation, which is easy to accomplish under the DDM.

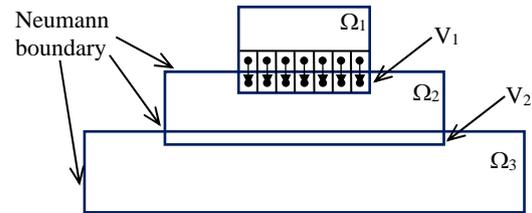


Fig. 3 Non-overlapping DDM for the pyramid-shape IC model.

Fig. 3 shows a non-overlapping DDM partition from the viewpoint of FVM and circuit equivalence. Taking a top-to-bottom iteration order as an example, the subdomains are solved in the order of Ω_1 , Ω_2 , and Ω_3 . To start, initial temperatures should be assumed on the bottom surfaces of Ω_1 and Ω_2 . And, the heat flow across the top surface of Ω_2 and Ω_3 derived from the solution of Ω_1 and Ω_2 respectively constitute the Neumann boundary conditions for solving Ω_2 and Ω_3 . The difference of temperatures on interfaces V_1 and V_2 should be checked to see if the iteration has reached a convergence. The convergence rate of DDM can be accelerated by a relaxed iterative scheme.

In [12], the details of the DDM for thermal simulation is presented, which assumes each subdomain is of rectangular shape and can be solved with the fast Poisson solver (FPS) [3] in $O(n \log n)$ time complexity. The experiments in [12] show that the DDM converges quickly (in 8 or 9 iterations). Compared to the state-of-the-art iterative equation solvers [3, 14], the DDM exhibits 10X memory saving and 2X or more speedup for larger cases.

The idea of solving subdomains with different discretization resolutions was also verified in [12]. By increasing the discretization steps of the subdomains of heat spreader and heat sink in turn, the results show that up

to 18X speedup is attained with less than 0.5% temperature error. Note that a linear interpolation is used to convert the quantities across the subdomain interface, which guarantees the good accuracy of chip temperatures.

The domain-decomposition approach enables fast thermal simulation with high-resolution chip discretization, which is favorable for the analysis and optimization of designs with complicated power profiles.

4. The Hybrid Random Walk Approach

4.1 Random walk method for thermal analysis

There are mainly two kinds of random walk method: the discrete random walk (DRW) method on an existing mesh grid, and the floating random walk (FRW) method [10, 11]. In the FRW method, the cubic or spherical transition domains with variable size are employed, and each step of walk is from its center to its boundary. The DRW has been applied to the problems of power grid analysis and thermal analysis [6-9], which is called *generic random walk* (GRW) method. In a game of GRW, a walker starts at a node in the grid, for which the voltage/temperature is calculated. The walker then randomly visits a neighbor node following the probability

$$P(i, j) = \frac{G_{ij}}{\sum_j^{d(i)} G_{ij}}, \quad (6)$$

where $d(i)$ is the edge degree of node i , and G_{ij} is the electric/thermal conductance between node i and its neighbor j . At each node, the walker receives a reward of

$$T_r(j) = \frac{p_j}{\sum_j^{d(i)} G_{ij}}, \quad (7)$$

where p_j is the current/power injected into current node j . The walk terminates if the walker hits a node with known voltage/temperature (home node), where the walker receives the final reward of the value of voltage or temperature. The total amount of money collected by the walker is an estimation of the voltage/temperature of the starting node. Due to the central limit theorem, the calculated node voltage/temperature obeys the normal probability distribution, whose variance is inversely proportional to the number of walks. This gives a tradeoff between runtime and accuracy.

If the GRW method is applied to the thermal simulation with the whole pyramid-shape IC model, the large simulation domain will make walks with a large number of steps. Also note that R_{amb} modeling the accurate convective boundary is larger than the thermal resistance of cells for several orders of magnitude. Therefore, they largely affect the computational speed of GRW method.

Fig. 4 shows a comparison of the GRW and FRW method. Clearly, a walk in GRW involves larger number of steps than a walk in FRW. However, the FRW method is suitable for the Laplace equation [10, 11], while for the thermal problem the power item $p(x, y, z)$ in (1) and

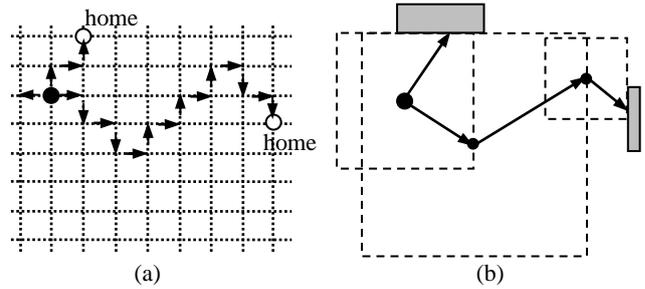


Fig. 4. Generic random walks on a mesh grid (a), and floating random walks using cubic transition domains (b).

complex boundary conditions bring difficulties to FRW method. So, it may be a good idea to combine GRW and FRW. In the smaller power-source region of the whole IC model, the GRW has to be used. The FRW applies to the other heat dissipation regions, which will contribute speedup to the overall random walk procedure.

4.2 Techniques for the hybrid random walk approach

The idea of hybrid random walk approach is illustrated in Fig. 5, where there is a walk starting from the GRW region and then behaving in the FRW manner. This hybrid random walk scheme avoids the difficulty of FRW for handling the power item in the thermal simulation.

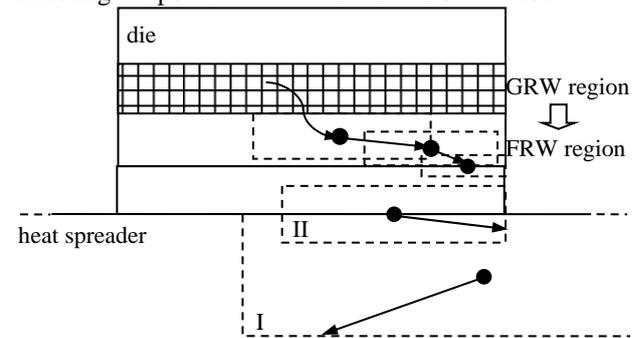


Fig. 5. A portion of the pyramid-shape IC model for the illustration of the hybrid random walk method. Transition domains with cuboid shape are shown.

The choice of transition domain affects the efficiency of FRW. In our problem, the random walk is performed in the pyramid-shape IC geometry, where the silicon die has much larger width than its height. The transition domain of cuboid shape, rather than the cubic domain for capacitance extraction [10], should be used to achieve better efficiency. This enables jumping far in lateral directions, and reduces the length of walk. To make the walk crossing the interface of different thermal materials, we also need the cuboid transition domain with two halves of different materials. These two kinds of transition domains are labeled with "I" and "II" in Fig. 5. Note that the FRW transition domain is bounded by the boundaries of the geometric model and GRW region. Near these boundaries, the FRW will be degraded to the GRW.

For an arbitrary transition domain where the Laplace equation of temperature $T(\mathbf{r})$ holds, we have:

$$T_c = \iiint_{\mathcal{V}} G_s(\mathbf{r}) T(\mathbf{r}) d\mathbf{r}. \quad (8)$$

Here, T_c is the temperature at the center of a transition domain surrounded by S . $G_s(r)$ is the surface Green's function, which has non-negative value and can be regarded as the probability density function (PDF). If we discretize S into small panels, with $G_s(r)$ we can calculate the transition probabilities to these panels, which makes the FRW feasible [10]. More importantly, these transition probabilities can be pre-calculated and recalled during the FRW procedure, which largely reduces the runtime.

To further reduce the time of performing a FRW hop, we propose to characterize the transition domain with a hop-target table (HTT), which records the target boundary points in a long random walk process. This HTT is pre-calculated with a GRW procedure. While performing the FRW, for each hop we randomly pick an integer k from $[0, M)$, where M is the length of HTT. Then, the k -th element of HTT indicates this hop's target. The tradeoff may be the memory usage of the array for destinations, which is affordable in our problem of thermal analysis.

The Neumann boundary and convective boundary can be handled in a more efficient way (otherwise GRW is used). Other than the path reflection technique [15], a Neumann-specific transition domain can be employed, which is the one with a sidewall being Neumann boundary. Using the similar pre-characterization procedure described, we can characterize the FRW transition behavior from the domain's center to its five non-Neumann boundary faces. For the convective boundary, more complex technique can be used, which considers (5) and guarantees the accuracy.

4.3 Numerical results

Two chip structures are used in experiments. Structure 1 is a four-core 2-D chip, i.e. the Testcase no. 2 in [3]. The transverse dimensions of the die, spreader, and sink are $1\text{cm} \times 1\text{cm}$, $3\text{cm} \times 3\text{cm}$ and $7\text{cm} \times 7\text{cm}$. Structure 2 is a chip imitating the IBM POWER6 microprocessor. Except that the die is of $1.6\text{cm} \times 2\text{cm}$, its transverse dimensions are the same as Structure 1. The power maps of the two chips are shown in Fig. 6 (each with total 175W). The heat convective coefficient at the surface of heat sink is $8700\text{ W}/(\text{K} \cdot \text{m}^2)$ [5]. The ambient temperature is set to 20°C .

From the test structures, with different resolution of discretization we obtained six test cases. For them, the node temperatures on the power-source layer are calculated with different random walk methods (see Table

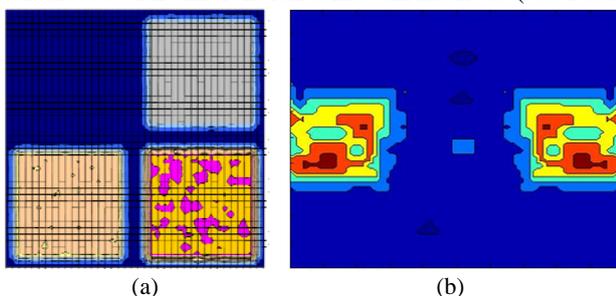


Fig. 6. The power maps for (a) Structure 1 and (b) Structure 2.

I). The hybrid0/2 are the basic HRW method and the one with special treatments of Neumann and convective boundaries, respectively. We see that the proposed techniques largely reduces the length of walk, and achieves up to 261X speedup. We have used the results got from Matlab “\” to validate the accuracy of HRW. For more details and experiments, please refer to [16].

Table I. The Average Runtime of Random Walk Methods for Calculating the Temperature of a Node (time in unit of second)

Test case	#node	GRW			Hybrid random walk			
		time	#walk	#hop	hybrid0	hybrid2	#hop	Sp*
1-1	5.24e5	49.8	5471	2.34e5	28.65	2.76	8.77e3	18
1-2	4.19e6	199	5522	9.28e5	48.2	4.12	8.13e3	48
1-3	6.55e7	949	6409	5.64e6	57.37	3.64	7.52e4	261
2-1	5.33e5	35.5	3576	2.52e5	32.86	1.26	1.05e4	28
2-2	4.26e6	143	3709	9.90e5	43.71	2.85	1.79e4	50
2-3	6.66e7	762	3281	5.98e6	72.7	5.17	3.12e4	147

5. Conclusions

A domain-decomposition method (DDM) and a hybrid random walk (HRW) method are presented for the thermal simulation of whole IC model with heat sink components. The both approaches are orders of magnitude faster than existing techniques, while preserving good accuracy. Compared to DDM, the HRW method is more general for handling complex IC thermal models.

REFERENCES

- [1] S.-C. Lin and K. Banerjee, “Thermal challenges of 3D ICs,” *Wafer-Level 3D ICs Process Technology*, Springer 2008.
- [2] S. Borkar, *Proc. DAC*, Jun. 2011, pp. 214.
- [3] H. Qian, S. S. Sapatnekar, and E. Kursun, *ACM Trans. Design Automation of Electronic Systems*, vol. 17, 2012, article 32.
- [4] P. Li, L. T. Pileggi, M. Asheghi, and R. Chandra, *IEEE Trans. Computer-Aided Design*, vol. 25, 2006, pp. 1763.
- [5] Y. Zhan and S. S. Sapatnekar, *IEEE Trans. Computer-Aided Design*, vol. 26, 2007, pp. 1661-1675.
- [6] H. Qian, S. R. Nassif, S. S. Sapatnekar, *IEEE Trans. Computer-Aided Design*, vol. 24, 2005, pp. 1204-1224.
- [7] T. Miyakawa, K. Yamanaga, H. Tsutsui, H. Ochi, and T. Sato, in *Proc. GLSVLSI*, 2011, pp. 211-216.
- [8] E. Wong and S. K. Lim, in *Proc. DATE*, 2006, pp. 878-883.
- [9] J. Guo, S. Dong, and S. Goto, in *Proc. ASICON*, Oct. 2009, pp. 771-774.
- [10] W. Yu, H. Zhuang, C. Zhang, G. Hu, Z. Liu, *IEEE Trans. Computer-Aided Design*, vol. 32, 2013, pp. 353-366.
- [11] C. Zhang and W. Yu, *IEEE Trans. Computer-Aided Design*, vol. 32, 2013, pp. 1633-1637.
- [12] W. Yu, T. Zhang, X. Yuan, and H. Qian, *IEEE Trans. Computer-Aided Design*, vol. 32, 2013, pp. 2014-2018.
- [13] K. Sun, Q. Zhou, K. Mohanram, and D. C. Sorensen, in *Proc. ICCAD*, 2007, pp. 54-59.
- [14] J. Yang, Z. Li, Y. Cai, and Q. Zhou, in *Proc. ICCAD*, Nov. 2011, pp. 482-487.
- [15] P. Maffezzoni and A. Brambilla, *IEEE Electron Devices Letters*, vol. 23, 2002, pp. 351-353.
- [16] Y. Liang, W. Yu, and H. Qian, in *Proc. ASP-DAC*, Jan. 2014, pp. 849-854.