Electrothermal Analysis and Optimization Techniques for Nanoscale Integrated Circuits

Yong Zhan, Brent Goplen, and Sachin S. Sapatnekar
Department of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455, USA.

Abstract— **With technology scaling, on-chip power densities are growing steadily, leading to the point where temperature has become an important consideration in the design of electrical circuits. This paper overviews several methods for the analysis and optimization of thermal effects in integrated circuits. Thermal analysis may be carried out efficiently through the use of finite difference methods, finite element methods, or Green function based methods, each of which provides different accuracy-computation tradeoffs, and the paper begins by surveying these. Next, we overview a restricted set of thermal optimization methods, specifically, placement techniques for thermal heat-spreading, and then we conclude by summarizing a set of future directions in electrothermal design.**

I. INTRODUCTION

Thermal considerations are playing an increasingly important role in high-performance integrated circuits, necessitating a greater role for the thermal analysis and optimization in the design cycle. Temperature effects are important for several reasons. First, they can cause the transistor delay to change: elevated temperatures cause the threshold voltage to drop and degrade the mobility: depending on which effect wins out, the delay can either increase or decrease. Second, wire resistances increase with temperature, leading to larger interconnect delays. Third, leakage power is particularly sensitive to thermal variations, and varies exponentially with temperature. Worse yet, such an increase in leakage power could, in turn, lead to an increase in the temperature, leading to the potential for positive feedback and an effect known as thermal runaway, in which the chip can physically burn out. Fourth, high temperatures can lead to a variety of reliability problems, including negative temperature bias instability (NBTI), oxide breakdown, and electromigration.

As a result, there has been a great deal of research in the area of thermal analysis and optimization in recent years. This paper overviews some of these issues. Section II presents methods for thermal analysis, and is followed by Section III, which outlines methods for thermally-driven optimization. We conclude with a description of future directions in Section IV.
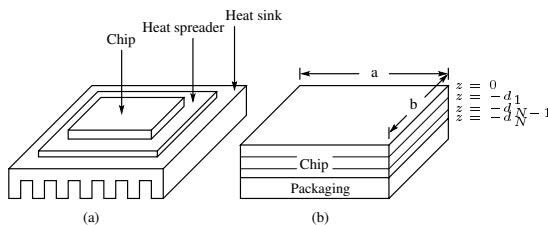
II. THERMAL ANALYSIS

Fig. 1. Schematic of a VLSI chip with packaging (a) integrated circuit and the packaging structure (b) simplified model of the chip and packaging.

Fig. 1(a) shows an integrated circuit chip with the associated packaging, and Fig. 1(b) shows a schematic of the structure in Fig. 1(a) where the packaging including the heat spreader and the heat sink has been simplified, but the multilayered structure of the

chip is explicitly shown. The steady state temperature distribution inside the chip is governed by Poisson's equation

$$\nabla^2 T(\mathbf{r}) = -\frac{g(\mathbf{r})}{k_{l(\mathbf{r})}} \tag{1}$$

where $\mathbf{r} = (x, y, z)$, $T(\mathbf{r})$ is the temperature (°C) distribution inside the chip, $g(\mathbf{r})$ is the volume power density (W/m^3), and $k_{l(\mathbf{r})}$ is the thermal conductivity (W/(m·°C)) of the layer where point $\mathbf{r}$ is located [1]. The vertical surfaces and the top surface of the chip are assumed to be adiabatic [2], and the bottom surface of the chip is assumed to be convective, with an effective heat transfer coefficient h (W/(m^2·°C)) [3]. In mathematical form, these boundary conditions can be expressed as

$$\left.\frac{\partial T(\mathbf{r})}{\partial x}\right|_{x=0,a} = \left.\frac{\partial T(\mathbf{r})}{\partial y}\right|_{y=0,b} = 0 \tag{2}$$

$$\left.\frac{\partial T(\mathbf{r})}{\partial z}\right|_{z=0} = 0 \tag{3}$$

$$k_N \left.\frac{\partial T(\mathbf{r})}{\partial z}\right|_{z=-d_N} = h(T(\mathbf{r})|_{z=-d_N} - T_a) \tag{4}$$

where T_a is the ambient temperature, and k_N is the thermal conductivity of the bottom layer of the chip. In addition, we enforce the continuity conditions at the interface between adjacent layers within the multilayered chip, i.e.,

$$T(\mathbf{r})|_{z=-d_i+\epsilon} = T(\mathbf{r})|_{z=-d_i-\epsilon} \tag{5}$$

$$k_i \left.\frac{\partial T(\mathbf{r})}{\partial z}\right|_{z=-d_i+\epsilon} = k_{i+1} \left.\frac{\partial T(\mathbf{r})}{\partial z}\right|_{z=-d_i-\epsilon} \tag{6}$$

where ϵ is an infinitesimally small quantity and k_i is the thermal conductivity of the ith material layer in the multilayered chip structure. Transient analysis is not explicitly treated here, but extensions of these methods as well as thermal-ADI techniques [4] may be used.

The most widely used numerical techniques for full-chip thermal analysis fall into three broad categories: finite difference, finite element and boundary element methods using Green functions. These are discussed in the following sections.

A. Finite Difference Analysis

The finite difference method (FDM) meshes the chip into (typically) rectangular cuboidal regions, assuming that the temperature of the region is represented by the temperature at its center. Writing the partial derivative as a finite difference between the temperature in neighboring regions leads to a well-established *thermal-electrical analogy*, where the thermal conductivity of the material is mapped to thermal resistors, and power sources (which correspond to the power drawn by logic cells) are mapped to current sources. This leads to a formulation of the type

$$G\mathbf{T} = \mathbf{P} \tag{7}$$

where G is the conductance matrix of thermal resistances, $\mathbf{T}$ is the vector of unknown temperatures at the center of each region, and $\mathbf{P}$ is the vector of power sources.

These equations closely resemble the equations for a power grid, and therefore, can be solved by a variety of techniques. Interestingly, although the substrate is also gridded into regions, it is often the case that the only points of interest are in the topmost device layer, and

therefore, the bulk can be macromodeled using Schur decomposition methods as proposed in [5], and improved upon in [6].

B. Finite Element Analysis

As in the case of the FDM, finite element analysis (FEA) also meshes the design space into elements. Various element shapes can be used such as tetrahedra and hexahedra, but a rectangular prism is ideal in that it makes the meshing of the rectangular-shaped substrate easier, and it can simulate heat conduction in lateral directions without aberrations in the prime directions.
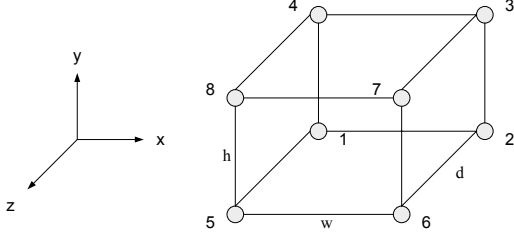

Fig. 2. An eight-node rectangular prism element for FEA.

In FEA, the temperatures are calculated at discrete points, the nodes of the elements, and the temperatures elsewhere within the elements are interpolated using a weighted average of the temperatures at the nodes. In deriving the finite element equations, the differential equation describing heat conduction is approximated within the elements using this interpolation. For an 8-node hexahedral element shown in Fig. 2, a trilinear interpolation function is used to describe the temperature within each element based on the nodal temperatures:

$$T(x, y, z) = \mathbf{N}^T \mathbf{T} \tag{8}$$

where $\mathbf{N} = [N_1 N_2 \cdots N_8]^T$ and $\mathbf{T} = [T_1 T_2 \cdots T_8]^T$, where T_i is the temperature at the ith of eight vertices of the rectangular prism, and N_i is the shape function for the ith vertex. The shape functions are determined by the width, w, height, h, and depth, d, of the element.

Similar to circuit simulation using the modified nodal analysis (MNA) method [7], stamps are created for each element and added to the global system of equations. In FEA, these stamps can be derived from the element interpolation functions using the variational method [8]. For an eight-node rectangular prism, a heat conduction stamp is produced as an 8×8 matrix, whose rows and columns correspond to the element's nodes. Stamps are also produced for surfaces exposed to convective boundary conditions. These element matrices are combined into a global matrix, K, by adding the matrix components that correspond to the same node in the global mesh together, and this produces a global system of equations

$$K\mathbf{T} = \mathbf{P} \tag{9}$$

where $\mathbf{T}$ is the vector of nodal temperatures and $\mathbf{P}$ is the vector of nodal powers. If isothermic boundary conditions are present, the resulting fixed temperatures and their corresponding rows and columns are removed from the system of equations with the right hand side modified accordingly.

C. Enhanced solution techniques for FDM and FEM

Several computational techniques may be employed to improve the solution speed of FDM and FEM. We overview two such methods in this section.

1) Random walk methods: Random walk methods have been used very successfully for the analysis of large RC networks, in the context of power grids [9], [10]. Such methods can easily be applied to the large resistive networks that appear in steady-state thermal analysis, and can be extended to transient thermal analysis as well. They can perform incremental analysis rapidly and efficiently, so that they are excellent candidates for incremental placement.

Consider the solution of a resistive network for the voltages; the thermal analog, of course, is that the voltages in the resistive network

are the temperatures, and the current sources are the power values. It has been shown that the solution to these networks can be obtained through a set of random walks on a graph. The nodes and edges of this graph are the nodes and connections in the initial circuit, respectively, and the transition probabilities from a node along each edge correspond to the ratio of the edge conductance to the total conductance connected to the node. A random walk begins at any node and along the way, at each node, it pays a levy that is related to the ratio of the current source connected to that node to the total conductance at the node. When the walk reaches a node whose voltage is known, it terminates and receives a reward that equals the voltage at the node. For each node, the expected value of earnings over all random walks beginning at that node can be shown to be the voltage of the node. A fuller exposition of this method is provided in [11], and it is shown to be efficient at finding the solution to a large network. For a more exact solution and better numerical efficiency, the method in [12] uses the approximate solution from the random walk solver to efficiently create a preconditioner for an iterative solver.

2) Multigrid Methods: An alternative method for solving systems of FDM equations may use the multigrid method, a multi-level iterative scheme. The essential idea of a multigrid scheme is that at every level, it coarsens the grid and then refines it: the coarsened grid is smaller and hence easier to solve, and captures the low-spatial-frequency component of thermal variations, while the refined grid capture high-spatial-frequency variations. Moving between a coarsened solution and a refined solution requires the use of interpolation and restriction operators, whose definitions are a key step in using multigrid methods. Depending on the coarsening/ refinement sequence, this may result in schemes such as the V cycle or the W cycle, or alternatives. Conventional multigrid schemes are classified as either geometric multigrid methods, which use some knowledge of the problem structure to perform the coarsening and refinement operations, or algebraic multigrid methods, which perform these steps automatically. The work in [13] uses a geometric multigrid method for solving FDM equations. A grid hierarchy is defined, with coarse grid operators

D. Green Function Based Methods

An alternative to the FEM and FDM methods, which mesh up the entire substrate, is a boundary element method using Green functions. The partial differential equation to be solved for thermal analysis is linear when the material properties are region-wise uniform, and therefore, conceptually, the problem can be solved by superposition, considering one source at a time. A Green function enables such a computation: it is the response in a field region to a power source in a source region; in the presence of multiple power sources, superposition can be used to sum up the responses at a field point due to each of the sources.

Let $G(\mathbf{r}, \mathbf{r}')$, with $\mathbf{r} = (x, y, z)$ and $\mathbf{r}' = (x', y', z')$, be the distribution of temperature above T_a in the multilayered chip structure when a unit point power source of 1W is placed at position $\mathbf{r}'$. Then $G(\mathbf{r}, \mathbf{r}')$ satisfies the equation

$$\nabla^2 G(\mathbf{r}, \mathbf{r}') = -\frac{\delta(\mathbf{r} - \mathbf{r}')}{k_{l(\mathbf{r})}} \tag{10}$$

and the boundary conditions

$$\left.\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial x}\right|_{x=0,a} = \left.\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial y}\right|_{y=0,b} = 0 \tag{11}$$

$$\left.\frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial z}\right|_{z=0} = 0 \tag{12}$$

$$\left.k_N \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial z}\right|_{z=-d_N} = hG(\mathbf{r}, \mathbf{r}')|_{z=-d_N} \tag{13}$$

$$G(\mathbf{r}, \mathbf{r}')|_{z=-d_i+\epsilon} = G(\mathbf{r}, \mathbf{r}')|_{z=-d_i-\epsilon} \tag{14}$$

$$\left.k_i \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial z}\right|_{z=-d_i+\epsilon} = \left.k_{i+1} \frac{\partial G(\mathbf{r}, \mathbf{r}')}{\partial z}\right|_{z=-d_i-\epsilon} \tag{15}$$

where $\delta(\mathbf{r}, \mathbf{r}') = \delta(x - x')\delta(y - y')\delta(z - z')$ is the three-dimensional Dirac delta function, and $G(\mathbf{r}, \mathbf{r}')$ is the Green function. The temperature field under an arbitrary power density distribution can be

obtained easily as

$$T(\mathbf{r}) = T_a + \int_0^a dx' \int_0^b dy' \int_{-d_N}^0 dz' G(\mathbf{r},\mathbf{r}')g(\mathbf{r}') \qquad (16)$$

For thermal problems encountered in chip design, both the source regions, where powers are generated, and the field regions, whose temperatures are to be computed, are located on discrete planes. Thus, in the following analysis, we will focus on a single source plane and a single field plane, i.e., a particular z and z'. For these planes, it can be shown [14] that the Green function is given by

$$G'(x, y, x', y') \triangleq G(\mathbf{r},\mathbf{r}')\,|_{z,z'} =$$
$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} C_{mn} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) \cos\left(\frac{m\pi x'}{a}\right) \cos\left(\frac{n\pi y'}{b}\right)$$
$$(17)$$

where the coefficient C_{mn} only depends on z and z'.

The above expression is complicated, both visually and computationally, and it involves a double summation to infinity. Fortunately, several methods are available for managing the computation. The work in [14], based on the substrate analysis methods in [15], uses the discrete cosine transform (DCT) and table lookups to accelerate the Green function based thermal analysis; for the analysis of a single layer, the computational complexity is $O(N_g^2)$, where N_g is the number of regions. An improved method in [16] reduces the complexity from quadratic to $O(N_g log N_g)$. The essential idea is to recognize that the bottleneck corresponds to a convolution operation, and this can be performed efficiently in the frequency domain: the primary cost here is in the transform from the space domain to the frequency domain.

III. Thermal Placement

While several methods can be used for thermal optimization, we will focus here on thermal placement.

At the placement stage, as far as the thermal constraints are concerned, the goals are to minimize the maximal on-chip temperature gradient and obtain an even temperature distribution, while controlling conventional metrics such as the wire length. Thermal analysis involves the solution of a system of linear equations of the type

$$G\mathbf{T} = \mathbf{P} \qquad (18)$$

where G is the thermal conductance matrix, $\mathbf{T}$ is the vector of temperatures, and $\mathbf{P}$ is the vector of power dissipations. For the finite difference method, G refers to the matrix associated with the thermal conductances, while for the finite element method, this corresponds to the stiffness matrix. The value of each P_i is not constant, but may change depending on which cell is located in a particular grid of the layout. Additionally, the power consumption of a cell varies with the interconnect capacitance that it drives, i.e., the length of the nets that it drives. During placement, these values are liable to change. The total power is actually dissipated by both the switching transistors and the interconnecting wires. Except for long global wires, the driver resistance is typically much larger than the metal resistance, and therefore most of the power is dissipated in the cells, and it is reasonable to ignore the part consumed by the metal wires. Even though self-heating of wires plays a very important role in the electromigration lifetime of the metal wires [17], during the placement stage, it may be ignored. It is potentially possible to take this into account during a later stage of placement using the congestion information.

A. Partitioning-based Placement

Partitioning-based approaches to placement are based on the idea of recursively dividing the layout into regions and assigning cells to each region. The key issue in partitioning-based placement, tackled in [6], is to simplify the thermal model at each level of partitioning to achieve the goal of placing the cells so that T is evenly distributed across the chip. The equation $G\mathbf{T} = \mathbf{P}$ cannot be directly used, since at each partition level, the only location information that is available is identity of the partitioning blocks that the cell belongs to. Assuming that all the cells belonging to a block are located at its center, the corresponding analysis will correspond to an incorrect thermal analysis of the partition. For example, it is easy to see that this will result in an exaggerated thermal gradient within the partition, since the temperature at the center of the partition will be much higher than that at its periphery.

For simplicity, consider the thermal model for a top-down bipartitioning process; this process can be easily extended to a top-down k-way partitioning process. At any one particular partition stage, a single block is being partitioned into two sub-blocks so that the cuts between the boundary of sub-blocks are minimized. For now, the actual temperature profile in the sub-blocks is not of direct concern, since cells inside the blocks will be further partitioned later, and this can be considered at that time. However, it is important to minimize the temperature discrepancy between the blocks. If some high-power cells are accumulated into one of the blocks, then at a later stage it will not be possible to move these cells out of the block, due to the divide-and-conquer nature of top-down partitioning. It is reasonable to assume that the temperature inside each of the sub-blocks are uniform, since if such an objective were to be enforced at every step of the partitioning, then a uniform temperature distribution would indeed result. Under this assumption, a simplified thermal model is obtained, and assumptions about the precise cell locations inside the sub-block need not be made.

In the first step in top down partitioning, the chip is partitioned into two blocks, the left block and the right block. For simplicity, assume that the number of thermal cells in each region is the same, although this assumption can easily be discarded. The thermal cells $i = 1 .. m/2$ will be said to be in the left block and cells $i = m/2 + 1 .. m$ in the right block. Now assume, as stated above, that the block on the left has an even temperature of T_l and the temperature of the block to the right is T_r. The thermal equation can now be simplified to:

$$\begin{bmatrix} \sum_{i=1}^{m/2} \sum_{j=1}^{m/2} G_{ij} & \sum_{i=1}^{m/2} \sum_{j=\frac{m}{2}+1}^{m} G_{ij} \\ \sum_{i=\frac{m}{2}+1}^{m} \sum_{j=1}^{m/2} G_{ij} & \sum_{i=\frac{m}{2}+1}^{m} \sum_{j=\frac{m}{2}+1}^{m} G_{ij} \end{bmatrix} \begin{bmatrix} T_l \\ T_r \end{bmatrix} =$$
$$\begin{bmatrix} \sum_{i=1}^{m/2} P_i \\ \sum_{i=\frac{m}{2}+1}^{m} P_i \end{bmatrix} \qquad (19)$$

This reasoning can be extended to a general case where the chip is partitioned into k regions, each with possibly a different number of cells, where a positive definite system of k equations in k variables must be solved.

The approach in [6] uses a top-down two-way partitioner based on the Fiduccia-Mattheyses algorithm [18], but can be extended naturally to incorporate a state-of-the-art multi-level partitioner. To reduce the computational cost for incremental temperature updates, the notion of an "effective thermal influence region" of a block is introduced. For a unit heat source on a block, this corresponds to the area outside of which the temperature induced by the unit heat source is less than a certain percentage of the maximum temperature induced by the unit heat source: this can be easily computed once the thermal resistance matrix is known. At each partitioning step, the cells are moved to provide a near-uniform thermal profile, while attempting to control the wire length.

B. Force-directed Placement

In force-directed methods, an analogy to Hooke's law is used by representing nets as springs and finding the placement corresponding to the system's minimum energy state. Attractive forces are created between interconnected cells and are made proportional to the separation distance and interconnectivity. Other design criteria such as cell overlap, timing, and congestion are used to derive the repulsive forces. After repulsive forces are added, the system is solved for the minimum energy state, i.e., the equilibrium location. Ideally, this minimizes the wire lengths while at the same time satisfying the other design criteria.

The work in [19] presents a force-directed approach to thermal placement. The application domain is in the design of 3D integrated circuits, where chips have multiple levels of active devices. Therefore, placement must be carried out in not just the xy-plane, but the entire xyz-space in three dimensions. In current technologies, in the z dimension, the number of layers is restricted to a small number. The work in [19] uses a force-directed framework with FEA-based thermal analysis, using repulsive forces to avoid hot spots. The thermal forces are calculated using the temperature gradient, which itself can be related to the stiffness matrix and its derivative. The temperature gradient determines both the direction and relative magnitude of the thermal forces, thereby moving cells away from areas with high temperature.

Fundamentally, force-directed methodologies involve minimizing an objective function corresponding to a summation of cost components from each net. For 3D layouts, this takes the form

$$c_{ij}\left[(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2\right] \quad (20)$$

where c_{ij} is the weight of the connection between the two nodes. If the c_{ij} coefficients are combined into a global C matrix, an objective function can be written for the entire system:

$$\frac{1}{2}\mathbf{x}^T C \mathbf{x} + \frac{1}{2}\mathbf{y}^T C \mathbf{y} + \frac{1}{2}\mathbf{z}^T C \mathbf{z} \quad (21)$$

where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are the x, y, and z coordinates of all cells and points of interest. This objective function can be minimized by solving the following three systems of equations:

$$C\mathbf{x} = \mathbf{f}_x, C\mathbf{y} = \mathbf{f}_y, C\mathbf{z} = \mathbf{f}_z \quad (22)$$

In the absence of external repulsive forces, the total force vectors, $\mathbf{f}_x$, $\mathbf{f}_y$, and $\mathbf{f}_z$, would be zero. The net stiffness matrix, C, describes the entire net connectivity. Fixed coordinate values, created by physical constraints such as I/O pads, can be used to reduce and solve the system of equations, much like the isothermic boundary conditions in FEA.

Generally, an iterative force-directed approach follows the following steps in the main loop. Initially, forces are updated based on the previous placement. Using these new forces, the cell positions are then calculated. These two steps of calculating forces and finding cell positions are repeated until the exit criteria are satisfied.

IV. Future Directions

Several issues remain to be solved in the area of electrothermal design:

- Leakage power is becoming a major contributor to the total power: leakage depends exponentially on temperature, and a rise in this component of power can itself cause the temperature to increase. Therefore, electrothermal analysis must be self-consistent in taking this feedback effect into account in the high-leakage regime. Since leakage power also varies significantly with process variations, even a circuit that is designed to meet its power and thermal requirements may fail these specifications after manufacturing. Adaptive methods such as the application of body biases may be used to allow a circuit to recover from the increased leakage power due to process and temperature variations [20], [21]. Moreover, temperature causes circuit delays to change, both by changing transistor characteristics and by altering interconnect delays, resulting in altered power-delay tradeoffs.

- A number of reliability issues come into play as temperatures rise. Negative temperature bias instability (NBTI) [22] causes the threshold voltages of devices to shift with time, and is an aging process that is accentuated under thermal stress. Electromigration is also hastened under high temperatures [23], [24], and effects such as these must be controlled.

- Traditional solution separate the problems of power reduction from the problem of heat removal. More efficient techniques can develop spot cooling methods for heat removal. One example of such a method utilizes thermal vias to remove heat from 3D integrated circuits [25].

Many of these problems have seen little research in the past, and are fertile grounds for future work.

References

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

[2] A. G. Kokkas. Thermal analysis of multi-layer structures. *IEEE Transactions on Electron Devices*, 21(11):674–681, November 1974.

[3] Y. K. Cheng, P. Raha, C. C. Teng, E. Rosenbaum, and S.-M. Kang. ILLIADS-T: An electrothermal timing simulator for temperature-sensitive reliability diagnosis of CMOS VLSI chips. *IEEE Transactions on Computer-Aided Design*, 17(8):668–681, August 1998.

[4] T.-Y. Wang and C. C.-P. Chen. 3-D thermal-ADI: A linear-time chip level transient thermal simulator. *IEEE Transactions on Computer-Aided Design*, 21(12):1434–1445, December 2002.

[5] C. Tsai and S. Kang. Cell-level placement for improving substrate thermal distribution. *IEEE Transactions on Computer-Aided Design*, 19(2):253–266, February 2000.

[6] G. Chen and S. S. Sapatnekar. Partition-driven standard cell thermal placement. In *Proceedings of the International Symposium on Physical Design*, pages 75–80, 2003.

[7] C. Ho, A. E. Ruehli, and P. Brennan. The modified nodal approach to network analysis. *IEEE Transactions on Circuits and Systems*, 22(6):504–509, June 1975.

[8] D. L. Logan. *A First Course in the Finite Element Method.* Brooks/Cole Pub. Co., 3 edition, 2002.

[9] H. Qian, S. R. Nassif, and S. S. Sapatnekar. Random walks in a supply network. In *Proceedings of the ACM/IEEE Design Automation Conference*, pages 93–98, 2003.

[10] H. Qian and S. S. Sapatnekar. Hierarchical random-walk algorithms for power grid analysis. In *Proceedings of the Asia/South Pacific Design Automation Conference*, pages 499–504, 2003.

[11] H. Qian, S. R. Nassif, and S. S. Sapatnekar. Power grid analysis using random walks. *IEEE Transactions on Computer-Aided Design*, 24(8):1204–1224, August 2005.

[12] H. Qian and S. S. Sapatnekar. A hybrid linear equation solver and its application in quadratic placement. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 2005.

[13] P. Li, L. T. Pileggi, M. Asheghi, and R. Chandra. Efficient full-chip thermal modeling and analysis. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 319–326, 2004.

[14] Y. Zhan and S. S. Sapatnekar. Fast computation of the temperature distribution in vlsi chips using the discrete cosine transform and table look-up. In *Proceedings of the Asia/South Pacific Design Automation Conference*, pages 87–92, 2005.

[15] R. Gharpurey and R. G. Meyer. Modeling and analysis of substrate coupling in integrated circuits. *IEEE Journal of Solid-State Circuits*, 17(4):305–315, March 1996.

[16] Y. Zhan and S. S. Sapatnekar. A high efficiency full-chip thermal simulation algorithm. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, 2005.

[17] K. Banerjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C. Hu. On thermal effects in deep sub-micron vlsi interconnects. In *Proceedings of the ACM/IEEE Design Automation Conference*, pages 885–891, 1999.

[18] C. M. Fiduccia and R. M. Mattheyses. A linear time heuristic for improving network partitions. In *Proceedings of the ACM/IEEE Design Automation Conference*, pages 175–181, 1982.

[19] B. Goplen and S. S. Sapatnekar. Efficient thermal placement of standard cells in 3d ics using a force directed approach. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 86–89, 2003.

[20] J. W. Tschanz, J.T. Kao, S. G. Narendra, R. Nair, D.A. Antoniadis, A.P. Chandrakasan, and V. De. *Adaptive Body Bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage. IEEE Journal of Solid-State Circuits*, 37(11):1396–1402, November 2002.

[21] S. Kumar, C. H. Kim, and S. S. Sapatnekar. Mathematically-assisted adaptive body bias (abb) for temperature compensation in gigascale lsi systems. In *Proceedings of the Asia/South Pacific Design Automation Conference*, 2005.

[22] D. K. Schroder and J. A. Babcock. Negative bias temperature instability: Road to cross in deep submicron semiconductor manufacturing. *Journal of Applied Physics*, 94:1–18, 2003.

[23] J. R. Black. Mass transport of aluminum by momentum exchange with conducting electrons. In *Proceedings of the IEEE International Reliability Physics Symposium*, pages 148–159, 1969.

[24] Z. Lu, W. Huang, J. Lach, M. Stan, and K. Skadron. Interconnect lifetime prediction under dynamic stress for reliability-aware design. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pages 327–334, 2004.

[25] B. Goplen and S. S. Sapatnekar. Thermal via placement in 3d ics. In *Proceedings of the International Symposium on Physical Design*, pages 167–174, 2005.