Thermal and Electromigration Analysis

by

Georgios Floros

Submitted to the Department of Electrical and Computer Engineering in partial fulfillment of the requirements for the degree of

Master in Science and Technology of Computer and Communication Engineering

at the

University of Thessaly

September 2015

Author:……………………………………………………………….................

Georgios Floros
Department of Electrical and Computer Engineering

Supervisors:……………………………………………………………..

Dr. Georgios Stamoulis
Professor of Electrical and Computer Engineering

Dr. Nestor Evmorfopoulos
Assistant Professor of Electrical and Computer Engineering

Dr. Ioannis Moondanos
Associate Professor of Electrical and Computer Engineering

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Abstract

Temperature consideration is one of the most critical challenges that come with technological evolution. The continuous effort for smaller sizes and greater performance as well as new 3D structure of integrated circuits together with the use of low-k dielectrics, cause non-negligible temperatures that has begun to outpace the ability of today's heat sinks to limit the on-chip temperature. The performance, reliability and power consumption of the devices are set under risk, while Joule heating effect worsens. Undoubtedly, thermal issues demand our concern and have to seriously been taken into account.
Moreover, electromigration is starting to be one of the most significant problems considering reliability in integrated circuits design. The problem is induced by the large current density and high temperature in circuit interconnections.
Furthermore, the continuous reduction of the size of the integrated and the simultaneous increase of the currents flowing in semiconductors have introduced challenges in the design that increasingly are taking into consideration electromigration.
In the present thesis is being introduced the methodology and implementation aspects of temperature and electromigration analysis tool.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Περίληψη

Η θερμική ανάλυση είναι μία από τις πιο κρίσιμες προκλήσεις που συνεπάγεται η τεχνολογική εξέλιξη. Η συνεχής προσπάθεια για μικρότερα μεγέθη και μεγαλύτερη απόδοση, καθώς και η νέα τρισδιάστατη δομή των ολοκληρωμένων κυκλωμάτων, μαζί με τη χρήση των low-k διηλεκτρικών, προκαλούν μη αμελητέα αύξηση της θερμοκρασίας η οποία δεν μπορεί να περιοριστεί από τον απαγωγέα θερμότητας.
Η απόδοση, αξιοπιστία και η κατανάλωση ισχύος των συσκευών τίθενται σε κίνδυνο, ενώ το φαινόμενο Joule heating, επιδεινώνεται. Αναμφίβολα επομένως, τα θερμικά ζητήματα απαιτούν την προσοχή μας και θα πρέπει να λαμβάνονται σοβαρά υπόψιν.
Επιπροσθέτως, η ηλεκτρομετανάστευση αρχίζει να γίνεται ένα από τα κύρια προβλήματα αξιοπιστίας στον σχεδιασμό ολοκληρωμένων κυκλωμάτων. Το πρόβλημα δημιουργείται από τη μεγάλη πυκνότητα του ρεύματος και τις ηψυλές θερμοκρασίες στις διασυνδέσεις του κυκλώματος. Επί πλέον, η συνεχόμενη μείωση του μεγέθους των ολοκληρωμένων και η ταυτόχρονη αύξηση των ρευμάτων που ρέουν στους ημιαγωγούς έχουν δημιουργήσει προκλήσεις στην σχεδίαση τους με παραμέτρους, λαμβάνοντας υπόψη την ηλεκτρομετανάστευση.
Στην παρούσα διπλωματική εργασία παρουσιάζεται η μεθοδολογία και η υλοποίηση ενός εργαλείου για θερμική ανάλυση και ανάλυση ηλεκτρομετανάστευσης.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Acknowledgements

With the completion of the present master theses I would like to express my sincere gratitude to my supervisors, Dr. George Stamoulis, Dr. Nestoras Evmorfopoulos and Dr. Ioannis Moondanos for their guidance and support during my postgraduate studies. Their essential observations made the accomplishment of this dissertation feasible and much easier.

Moreover, special thanks goes to all my friend and colleagues for their critical help during this thesis.

Finally, I would like to thank my family for their constant support and patience throughout all these years.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

To everyone who supported me

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Contents

1. Introduction 9
1.1 Thermal Problem 9
1.2 Electromigration Problem 11
1.3 Impact of the Theses 12
1.4 Structure of the Theses 12
2. Full Chip Thermal Analysis 13
2.1 The Thermal PDE 13
2.2 Finite Element Method 13
2.3 Finite Difference Method 15
2.4 Proposed Methodology Flow 17
2.5 Solution Method 18
3. Temperature Dependent Electromigration 21
3.1 Electromigration Physics 21
3.2 Simulation Modeling 21
3.3 Probabilistic Simulation 22
3.4 Proposed Methodology Flow 22
3.5 Solution Method 23
Figure VI Electromigration Workflow 24
4. Benchmarks and Results 25
4.1 DEF/LEF Introduction 25
4.2 Benchmarks 25
4.3 Results 26
5. Conclusion and Future Work 29
6. References 31

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

List of Figures

I. A schematic of a 3D integrated circuit……………………………………….......11
II. Discretization of a single layer…………………………………………………...15
III. Equivalent circuit of each node……………………………………………...16
IV. A complete 3D Network…………………………………………………...17
V. Final Circuit…………………………………………………………………...18
VI. Electromigration Workflow…………………………………………………24

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

List of Tables

I. ISCAS Benchmarks……………………………………………………………….25

II. Benchmark Explanation…………………………………………………………...27

III. Simulation Results………………………………………………………………28

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

1. Introduction

Power dissipation in integrated circuits has begun to outpace the ability of today's heat sinks to limit the on-chip temperature. While the formulation problems, by applying a thermal equivalent circuit, is prevalent and can be easily constructed, the corresponding 3D equations network has an undesirably time consuming solution. As a result, an efficient full chip thermal simulation is among the most challenging problems facing the EDA industry today. This work presents a new approach for fast thermal simulation of 3D ICs found in contemporary nanometer-scale designs.

1.1 Thermal Problem

Since the appearance of the first technological achievements, speed and performance have been the main targets inspiring inventors to evolve their previous attainments. More specifically, in semiconductor industry, there is a relentless effort for higher CMOS performance and functionality, greatly pushed by the customer needs and the competition between manufacturers. Undoubtedly, the electronics industry plays a leading role in economic, social and political development throughout the world. Therefore, the attempt for rises in integrated circuits integration density and speed will continue to keep a dominant position in experts' minds.
Temperature consideration is one of the most critical challenges that come with technological evolution. The continuous effort for smaller sizes and greater performance as well as the new 3D structure of integrated circuits has begun to outpace the ability of today's heat sinks to limit the on-chip temperature. Therefore full chip thermal analysis problems has drawn considerable attention in the past two decades.
Thermal analysis is important in ensuring the accuracy of timing, noise, and reliability analyses during chip design. The origin of thermal problems is in the fact that electronic circuitry dissipates power. This power dissipated on-chip is occurred in the form of heat, which, in a modern designed system, flows toward a heat sink. The power generated per unit area is often referred to as the heat flux. Temperature and power (or heat flux) are intimately related, but it is important to note that they are distinct from each other. For example the locations of the power sources in respect to the heat sink plays a major role in determining the on chip temperature.
However, power dissipation in integrated circuits has begun to outpace the ability of today's heat sinks to limit the on-chip temperature. As a result, thermal issues have come to the forefront, and thermally aware design techniques are likely to play a major role in the future. Why, though, is it important to do thermal analysis? There are two main reasons why thermal analysis is really necessary; firstly, because of the fact that temperature affects the performance and secondly, because both temperature and power are tightly correlated.
A significant issue accompanying the deeper entrance in nanometer sizes is the increased power density. The latter leads in elevated on chip temperature which puts the desired performance under risk, menacing the proper functionality of the devices. In addition to this, as chips warm up in a nonuniform way, local hot spots and spatial gradients are generated, with higher power densities and consequently higher local temperatures.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

The thermal menace also worsens due to the nowadays used multilayer 3D stacking, as well as the use of low-k dielectrics (Fig. I).Stacking multiple layers in a 3D volume promises density and performance. However, it requires extensive thermal consideration as the power density and temperature of these architectures can be quite high. In this case we need cooling a volume consisting of units placed on top of each other, instead of just cooling a planar surface. In addition to this, deep submicron technologies such as low-k dielectric materials have been introduced in order to achieve better performance, reducing interconnect capacitance and thus delay as well as cross talk noise. However, due to their lower thermal conductivity they are more susceptible to thermal effects.

Therefore, how to effectively analyze the chip-level three dimensional (3-D) thermal distribution and hot-spot locations is important. Several approaches have been proposed to perform thermal analysis. There are several methods that are used for thermal analysis modeling, such as FEM (Finite Element Method), analysis based on Green function and FDM (Finite Difference Method); hence, FDM is the most generic method that can handle complicated on-chip geometries such as nonuniform wiring structures.

Temperature increase is an inevitable aspect of the continuous scaling trend. Nevertheless, high temperature has significant impacts on chip performance and reliability. Leading to slower transistor speed, more leakage power consumption, higher interconnect resistance and reduced reliability, thermal issues constitute a major challenge that has to be seriously taken into consideration. Management of them remains key factor for future microprocessors and ICs. Great proof of their significance is the extensive existing research on thermal analysis models [2] [7] [11] [12] [19] [20] [22] [23], while many research group keep working on this subject trying to present more accurate thermal simulators for VLSI designs.

In this work, we adopt the FDM method with a state of the art Fast Transform Solver that to the best of our knowledge outperforms any existing solver in the case of the thermal equivalent network. The benefits of the proposed method are twofold: i) the proposed preconditioning mechanism can accelerate the convergence rate of the iterative solution method by greatly reducing the required number of iterations, and ii) from a computational point of view, it exhibits near-optimal computational complexity, low memory requirements, and great potential for parallelism, which can harness the computational power of parallel architectures, such as multicore processors or GPUs, thus further reducing the amount of time required for simulation.

Experimental results demonstrate that our method achieves a speedup 26.93 for a 10M node design over a state-of-the-art iterative method when GPUs are utilized. To the best of our knowledge, this is the first research approach that presents an algorithm for combined full-chip thermal simulation on parallel architectures.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

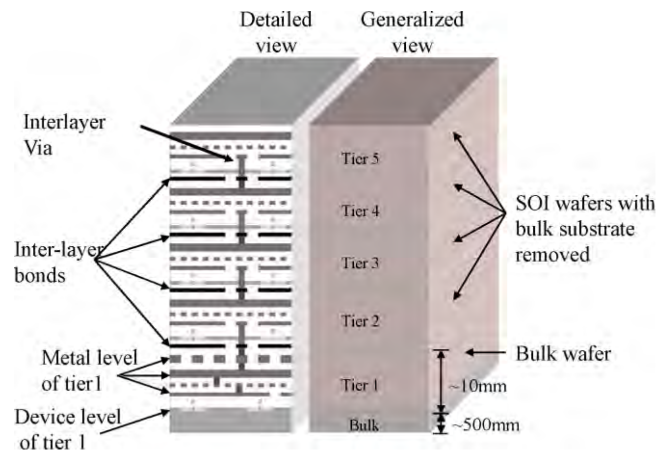


Figure I A schematic of a 3D integrated circuit.

1.2 Electromigration Problem

Electromigration has been observed since more than a hundred years. Despite this, the most significant modeling of the phenomenon was produced empirically by Black [3] [4] in 1966:

$$\mathrm{MTTF} = \mathrm{AJ}^{-2}\exp\left(\frac{Ea}{KT}\right) \qquad (1)$$

where *MTTF* is mean time to failure, *A* is a constant depends on interconnect length and material is the current density, E_a is the activation energy (e.g., 0.6 eV for aluminum), *k* is the Boltzmann constant and, finally, *T* is the operating temperature. Besides the primary work that has been done by Black, recent studies have attempted to characterize in more detail the effects of electromigration but still Black's equation is the most widely used.
Electromigration can result in open circuits and shorts which cause the catastrophic failure of the circuit. However, electromigration, in modern sub-45nm in which process variability affects the shape and the electrical parameters of interconnects, can produce circuit failures due to higher voltage drop (i.e. higher than expected resistance values) long before the catastrophic open- or short-circuits impede correct operation of the IC. Models and tools have been developed to produce relatively accurate reliability assessments of interconnects in modern integrated circuits during the design and the layout process, in order to provide feedback for reliability issues before fabrication. From this point of view several tools like BERT (BErkeley Reliability Tool) [9], ITEM [17], ERNI (Electromigration Reliability of Networked Interconnects), and ERNI-3D [1] [8] have been proposed in the academia. BERT and ITEM compute reliability issues taking as an input a given layout while ERNI –ERNI-3D are based on a hierarchical methodology based on filtering of immortal interconnect trees. All the above tools are taking into account mainly the maximum dc current and self-healing effect in order to provide estimations.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

1.3 Impact of the Theses

The purpose of the present theses is to provide a theoretical background about thermal and electromigration simulation and modeling in modern 3D IC structures.
From our point of view, early planning reliability tools must be integrated in every design flow and research community must pay attention in the increasing necessity of VLSI design with more reliability parameters. Moreover, in this work several novel techniques were used in order to achieve more accurate (in the Electromigration analysis) and faster results (in the case of full chip thermal analysis).

1.4 Structure of the Theses

In the first Chapter and introduction about the problem and the importance of accurate solutions was mentioned. Subsequently, Chapter 2 describes the full chip thermal analysis problem, giving a brief overview of the theory and the proposed methodology flow that we choose to implement. Similarly, Chapter 3 introduce the reader to some basic electromigration physics and the simulation approach that was implemented is analyzed. Finally, in the last two Chapter we present several results for both thermal and electromigration analysis (Chapter 4) and some conclusions with possible future implementations (Chapter 5).

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

2. Full Chip Thermal Analysis

In accordance with traditional thermal analysis, they use numerical approximations i.e. the finite difference, the finite element and the computational fluid dynamics. The calculations are intensive; especially for large scale 3D ICs and they demand the solution of linear equations that are given by the equivalent thermal circuit. Furthermore, the fast solvers are crucial in transient thermal simulation.

2.1 The Thermal PDE

There are three modes of heat transfer: conduction, convection and radiation. As we are interested in heat transfer in solids, we focus on the equation of heat conduction according to Fourier's law [15]:

$$q = -k_t \nabla T \qquad (2)$$

which means that the flow of heat at a point per unit area and per unit time, is proportional to the temperature gradient at that point and the heat flows in the direction of decreasing temperatures. In equation (2), q is the rate of heat flow, k_t is the thermal conductivity of the material and T the temperature. Analyzing the above equation we conclude in the following PDE relation:

$$\nabla q = -k_t \nabla^2 T = g(r,T) - \rho C_P \frac{\partial T(r,t)}{\partial t} \Rightarrow$$

$$\rho C_P \frac{\partial T(r,t)}{\partial t} = k_t \nabla^2 T(r,t) + g(r,t) \qquad (3)$$

Subject to the boundary condition

$$k_t(r,T)\frac{\partial T(r,t)}{\partial n_i} + h_i(r,t) = f_i(r,t)$$

where t is the time, g is the power density, C_P is the heat capacity of the chip material, ρ is the density of the material and r is the spatial coordinate of the point at which the temperature is being calculated, $n_i, f_i \, and \, h_i$ are the outward direction normal to the boundary surface i, an arbitrary function at the surface i and the heat transfer coefficient respectively.

2.2 Finite Element Method

According to the Finite Element Method (FEM), the design space is first discretized/meshed into elements such as tetrahedra or hexahedra.
Temperatures are then calculated at the nodes of the element, while temperatures elsewhere within the element are interpolated using the following function (for a hexahedral element):

$$T(x,y,z) = \sum_{i=0}^{8} N_i(x,y,z)Ti \qquad (4)$$

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

where Ti the temperature at node i and Ni the shape function for node i. If (xc, yc, zc) the center of the element and w, d, h the width, height and depth of the element respectively, Ni(x, y, z) can be written as:

$$N_i(\mathrm{x,y,z}) =$$

$$(\frac{1}{2}+\frac{2(x_i-x_c)}{w^2}(x-x_c))\mathrm{x}(\frac{1}{2}+\frac{2(y_i-y_c)}{d^2}(y-y_c))\mathrm{x}$$

$$(\frac{1}{2}+\frac{2(z_i-z_c)}{h^2}(z-z_c))$$

Using equation (4), the thermal gradient g can be calculated as:

$$g = \begin{bmatrix} \frac{\partial T}{\partial x} \\ \frac{\partial T}{\partial y} \\ \frac{\partial T}{\partial Z} \end{bmatrix} = BT$$

Where B:

$$\begin{bmatrix} \frac{\partial N_1}{\partial x} & \frac{\partial N_2}{\partial x} & \cdots & \frac{\partial N_8}{\partial x} \\ \frac{\partial N_1}{\partial y} & \frac{\partial N_2}{\partial y} & \ddots & \frac{\partial N_8}{\partial y} \\ \frac{\partial N_1}{\partial z} & \frac{\partial N_2}{\partial z} & \cdots & \frac{\partial N_8}{\partial z} \end{bmatrix}$$

Subsequently, stamps (in FEM they are called stiffness matrices K) are created for each element, using the variation method for an arbitrary element type:

$$K = \iiint_V B^T DBdV$$

where V the volume of the element and D = $\begin{bmatrix} k_{t,x} & 0 & 0 \\ 0 & k_{t,y} & 0 \\ 0 & 0 & k_{t,z} \end{bmatrix}$ with $k_{t,x}$ € x, y, z the thermal conductivities along x, y, z axis.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

According the boundary condition case (convective, conductive etc.), these stamps are accordingly calculated and together with the stamps from various 13 elements they are superposed to obtain the global stiffness matrix Kg in order to be incorporated to the global set of equations:

$$K_g \mathrm{T} = \mathrm{P} \qquad (5)$$

where T the vector of all the unknown temperatures and P the vector of power at the corresponding node.

2.3 Finite Difference Method

The heat conduction problem can be solved numerically using the Finite Difference Method (FDM). By applying the FDM method to the spatial derivative in the equation (2) we can conclude to the above formulation of the problem. If the chip is discretized along the 3 Cartesian Coordinated with length Δx, Δy and Δz respectively to generate a thermal grid as shown in figure 1 and the temperature of each node in the

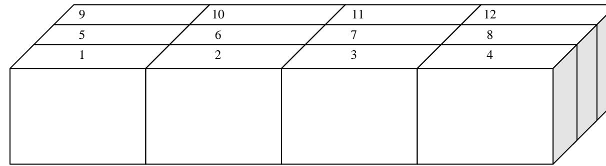


Figure II Discretization of a single layer

grid is represented by its location $T_{i,j,k}$, then the finite difference approximation for each point (i,j,k) can be expressed as:

$$\begin{aligned} C_p \Delta x \Delta y \Delta z \frac{dT}{dt} - k_t \frac{T_{i+1,j,k} - 2T_{i,j,k} + T_{i-1,j,k}}{\Delta x^2} \\ -k_t \frac{T_{i,j+1,k} - 2T_{i,j,k} + T_{i,j-1,k}}{\Delta y^2} \\ -k_t \frac{T_{i,j,k+1} - 2T_{i,j,k} + T_{i,j,k-1}}{\Delta z^2} \\ = \Delta x \Delta y \Delta z g(r,t) \end{aligned}$$

From this point, it can be easily derived the equivalent electrical model for the heat conduction problem with the temperature represented as voltage and the heat flow as electric current. The first term of the left hand side is represented as a capacitor and the rest terms in the left hand side are represented as a conductances as shown in figure 2. More specifically, the numerically value of the conductances can be derived from the above equations:

$$G_{W/E} = \frac{k_t \Delta y \Delta z}{\Delta x}$$

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

$$G_{S/N} = \frac{k_t \Delta x \Delta z}{\Delta y}$$

$$G_{TOP/DOWN} = \frac{k_t \Delta x \Delta y}{\Delta y}$$

and the value of the capacitor can be calculated as described in above equation:

$$C = \Delta x \Delta y \Delta z \rho C_P$$

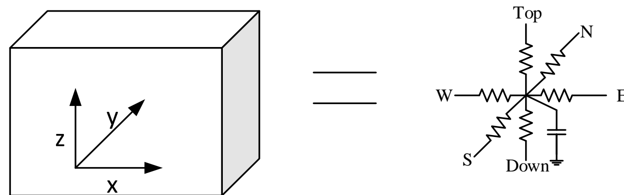


Figure III Equivalent circuit of each node

Finally, heat sources are modeled as current sources with value g(r,T)ΔxΔyΔz and connected to equivalent circuit nodes wherever exists heat dissipation.
These formulation concludes to the following system of ordinary differential equation:

$$\tilde{G}t(t) + \tilde{C}\dot{t}(t) = e(t)$$

where t(t) is the vector of all temperatures as a function of time, which are ordering with the convention of figure 3, $\tilde{G}$ is a symmetric tri-diagonal conductance matrix, $\tilde{C}$ is a diagonal matrix of cell capacitances and e(t) is a vector of heat sources as a function of time. This method can be easily extended in structures containing multiple layers and can accurate model even heterogeneous structures that can be found in modern chips (i.e. heat sinks).

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

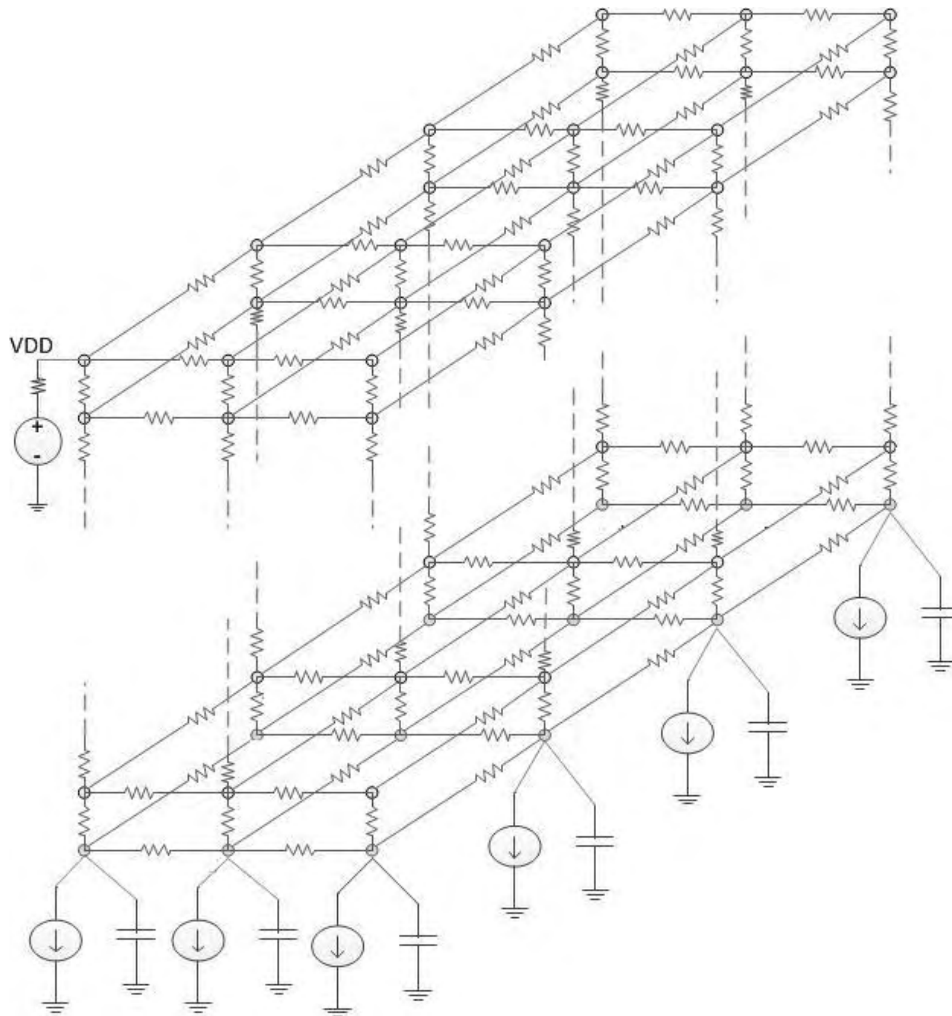


Figure IV A complete 3D Network

2.4 Proposed Methodology Flow

The purpose of this thesis is to investigate the influence of interconnections on temperature rise. Therefore, we begin examining LEF/DEF files and we parse the parts of the files that interest us. Considering the die area as a 3D grid of 1x1 micro we calculate the percentage of the block that each interconnect/via covers. As a next step, we model each block as 6 resistances forming a general network of resistances. Finally, using Finite Differences Method, we calculate a matrix containing the necessary coefficients that we then insert to a parallel fast preconditioner.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

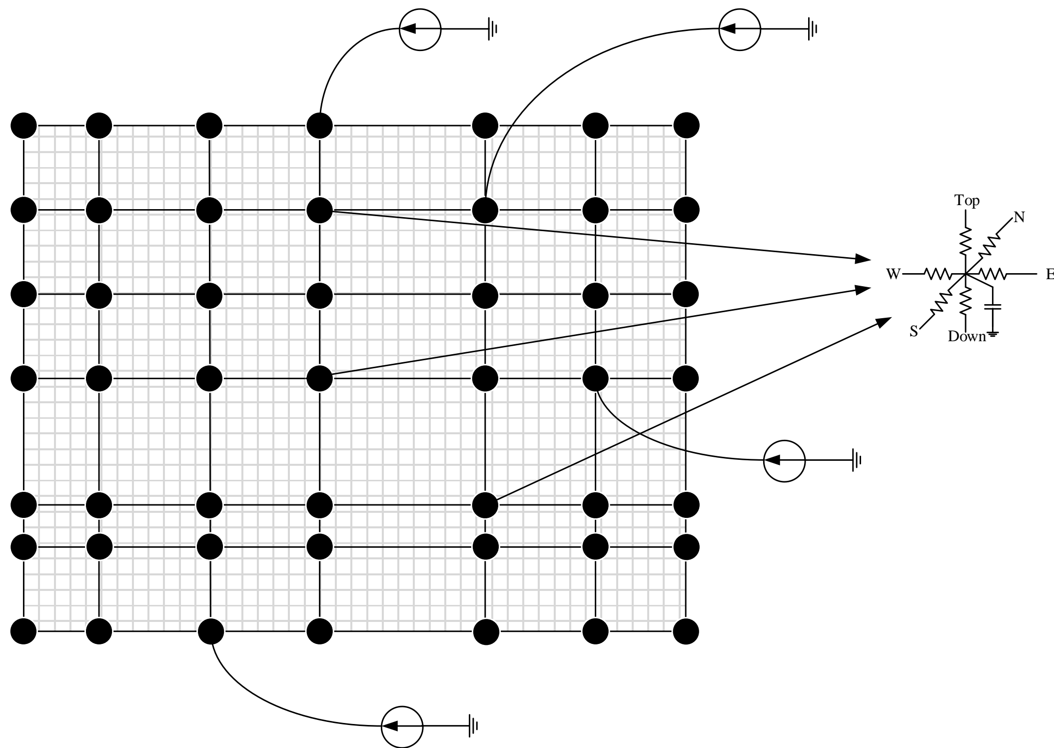


Figure V Final Circuit

2.5 Solution Method

In other words, the solution to the thermal analysis problem using FDM amounts to the solution of a circuit of linear resistors and current sources.
The ground node, or reference, for the circuit corresponds to a constant temperature node, which is typically the ambient temperature. If isothermal boundary conditions are to be used, this simply implies that the node(s) connected to the ambient correspond to the ground node.
The overall equations for the circuit may be formulated using modified nodal analysis and we may obtain a set of equations:

$$GT = P$$

Here G is an nxn matrix and T, P are n-vectors, where n corresponds to the number of nodes in the circuit. It is easy to verify that the G matrix is a sparse conductance matrix that has a banded structure, is symmetric and is diagonally dominant.
The finite difference approach is utilized for on-chip thermal analysis and it is known that a large number of nodes in the temperature vector can be eliminated.
There are several methods that are used for thermal analysis modeling, such as FEM (Finite Element Method), analysis based on Green function and FDM (Finite Difference Method); hence, FDM is the most generic method that can handle complicated on-chip geometries such as nonuniform wiring structures. Therefore, it can achieve very high accuracy in thermal analysis.
However, the direct application of this method usually involves meshing the entire substrate, which may lead to large problem sizes and relatively long runtimes. Using

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

the macromodeling techniques such as that presented in several scientific papers, it is possible to abstract away the nodes that the user of the thermal simulator is not interested in, and therefore, reduce the problem sizes in the finite difference analysis.
Additionally, building a macromodel still involves considerable effort. Therefore, the macromodeling approach is most effective under the situation where the chip geometry does not change but the thermal analysis needs to be performed multiple times, such as in the fixed-die thermal aware floorplanning and placement, because the time it takes to build the macromodel can be amortized.
Walking through the solution of the linear equations, there are not only direct but also iterative methods. Beginning with the direct ones, they were widely used in the past mostly because of their computational power; hence, they are forbidden simply because of their great computational cost. On the other hand, iterative methods include only interior products and matrix-vectors products and they consist a satisfying solution especially for sparse linear systems.
In accordance with preconditioners, they are used in order to accelerate solving linear systems. The purpose of using a preconditioner conditions that it converts a given problem into a form that is more suitable for numerical solution. Preconditioned problems are usually solved subsequently by using an iterative method.
Although there are many preconditioners that have been developed for general purposes, preconditioners cannot improve convergence as [24] points. The preconditioning of iterative methods reduces to a preconditioner solve step Mz = r in every iteration of the method, and effectively modifies the algorithm to solve the system $\mathrm{M}^{-1}Ax = \mathrm{M}^{-1}b$, which has the same solution as the original one $Ax = b$. If the preconditioner M approximates A in some way, then the condition number of the modified system $\mathrm{M}^{-1}Ax = M^{-1}b$ becomes very small and the iterative method converges very quickly.
Preconditioning is important in order to find a matrix M with the following properties: 1) the convergence rate of the preconditioned system $\mathrm{M}^{-1}Ax = \mathrm{M}^{-1}b$ is fast, and 2) a linear system involving M (i.e. $Mz = r$) - which effectively receives the whole burden of the algorithm it is solved much more efficiently than the original system involving A (i.e. $Ax = b$). The effect of good preconditioning is even more pronounced if the operations for the solution of $Mz = r$ can be performed in parallel,by applying a Fast Transform in a near-optimal number of operations in a sequential implementation.
Solving linear systems, not only FDM but also FEM lead to the formulation of problems that demand the solution of large linear equation systems. The matrixes that can describe these kinds of equations are usually sparse and positive defined.
Direct methods use variations of Gaussian elimination such as from LU factorization to first-factor matrixes in order to solve these equations. The system is solved using either forward or backward substitution.
The cost of the LU factorization is $O(n^3)$ for a dense nxn matrix but it becomes quite hyper linear for sparse systems. This step is followed by forward or backward substitution as it was mentioned before which costs O(n) for sparse systems, where the number of registrations per row is limited by a constant.
In case that a system is going to be valued for a large number of right-side vectors, that corresponds to different powers of vectors, LU factorization needs to be applied only once and its cost can be amortized solving multiple input vectors.
Studying iterative methods, it can be concluded that they are really efficient for sparse and positive defined matrixes. Such kind of iterative methods are following: Gauss-Jacobi, Gauss-Seidel and successive overrelaxation, as modern approximations based on Conjugate-gradient method or GMRES.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

The idea of iterative methods is to start with an initial solution and to redefine it successively, since there is convergence. Under those circumstances, it is possible to ensure convergence. Especially FDM matrixes have this ability.
Taking into consideration the above ideas we conclude that iterative methods are the appropriate method for the produced matrix of the full chip thermal analysis problem. As a result, we use the CG method with Fast Transform Precoditioner. This Precoditioner is dedicated for the nature of this resulting matrices. A complete study for this method can be found here[24].

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

3. Temperature Dependent Electromigration

3.1 Electromigration Physics

In this section, the focus shifts to the physical design of integrated circuits and the parameters that affect electromigration. According to Moore's law, the size of transistors becomes half every three years. This creates a greater need for power consumption as interconnects do not scale by the same factor in contrast to the area of the cross section which decreases continuously. Thus, the principal design parameters that have to be taken into account are the material, the size, the length, and the temperature of the interconnections. An important milestone in the production of reliable circuits was the replacement of aluminum by copper. Copper as a material can withstand about five times higher amperage than aluminum. This is mainly due to the higher activation energy caused by superior electrical and thermal conductivity and the higher melting point. Another parameter that greatly affects Black's equation is temperature. As shown in equation (1) the effect of temperature appears to be exponential, impacting significantly the final result. Current density, although an important factor, has lower impact. For example if we increase the temperature from 25° to 125° then we should reduce the current density of about 90%. Yet its magnitude, and in sub-45nm processes, its variation with time, play a significant role. The cross-section of the interconnection wires is also of paramount importance. Especially, in the era of extreme size reduction of the interconnection which leads to the significant increase of the current density. This has accumulative effect as temperature is increased at interconnection due to increased Joule heating. But this is not the only problem caused by the reduction of size of the interconnection. It may seem that the greater the width of the interconnect, the smaller the current density. This appears to have a positive effect on the MTTF of the integrated circuit. However, this is not always the case. There are cases where even though the process imposes a large reduction of the width of the interconnects, resistance to electromigration is increased. This contradiction is caused by the position of the grain boundaries of the integrated circuit due to the bamboo structure phenomenon. The last of the electromigration affecting parameters is interconnect length and the angle between the wires or contacts. A large cable has usually more angles and this reduces the factor A described in Black's equation due to current crowding at the corners. So it is more beneficial to have small wires that consist of straight lines. The same holds true for the contacts of the circuit where current crowding has to be dealt with new process techniques.

3.2 Simulation Modeling

In this section the proposed algorithm is presented in detail. The algorithm approaches the problem in two ways, a) through stochastic process in order to obtain the mean current and the maximum current and b) be also taking into consideration the self-healing effect, in the case that the current flowing through the interconnect a pulse or even a bi-directional current. The case of the maximum current is usually applied to the interconnects that have a unidirectional current flowing through them. As a result

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

the peak value of the current is calculated. This is a simple simulation obtaining the worst case, but it can give fast and relatively accurate results. This is due to the large value of MTTF and in some cases of small circuits we can have accurate results since the circuit can last more than 20 years. Focusing on the unidirectional currents, in the case of large circuits with power hungry power grids we need more accurate results. This leads to necessity of a more accurate algorithm.

3.3 Probabilistic Simulation

The breakthrough proposed by this work is to compute the mean current through a probabilistic scenario, taking advantage of the fact that we have a stationary stochastic process. This result is derived from the Borovkov theorem [5] and one similar idea, was implemented in [16]. Since the stochastic process is stationary it is quite easy to compute the average current of design. So taking into consideration this value we can have a more realistic calculation and, in our view, a more accurate one. Moreover, in bidirectional currents we compute, as described above, the average current of both direction and we apply a correction factor in order to take into account the self-healing effect. The validity of this method is shown here [18]. As a result the current that assigned to Black's equation is:

$$I = I_{go} - (0.8 \text{ or } 0.9) I_{come} \qquad (2)$$

3.4 Proposed Methodology Flow

The tool runs over all interconnect lines in order to determine if the interconnect can withstand the current density that runs through it. In order to obtain these results we only need to calculate the maximum current. The pseudo code implementing the algorithm is presented in order to clarify the flow. It should be mentioned that the input to the tool is the power grid or the extraction file from the given circuit along with the current waveforms, the specification file (i.e. LEF file) and the specific design parameters (i.e. maximum or average current mode, or pulse current).

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Electromigration algorithm

```
1:Input: Extraction/power grid (spice like), current waveforms,
physical specification
2:Output: Violations due to electromigration,MTTF
3:Mna=build_mna_system(extraction,current waveforms)
4:nodeVoltages=perform_transient_analysis (mna);
5:interconnectCurrentDensity= compute_current_density(nodeVoltages);
6:violation=check_for_violation (interconnectCurrentDensity,physical
Specifications);
7:if(violations>1)then
8:    return(Violation_Problem);
9: else
10:If(max_current_approach) then
11:   maxCurrent= get_max_current(interconnectCurrentDensity);
12:   MTTF_statistics= compute_MTTF(maxCurrent, physical
specifications);
13:   return(OK);
14:else if(mean_current_approach)
15:   meanCurrent=
compute_mean_current(interconnectCurrentDensity);
16:   MTTF_statistics=  compute_MTTF(meanCurrent,physical
specifications);
17:   return(OK);
18:else
19:   meanCurrentGo=
compute_mean_current(interconnectCurrentDensity);
20:   meanCurrentCome=
compute_mean_current(interconnectCurrentDensityCome);
21:   meanCurrent= meanCurrentGo -0.9* meanCurrentCome;
22:   MTTF_statistics= compute_MTTF(meanCurrent,physical
specifications);
23:   return(OK);
24:End
```

3.5 Solution Method

More specifically, the inputs to the tool are the extracted circuit netlist along with the specific parameters such as temperature, activation energy, and the currents drawn by each gate of the circuit, expressed in current sources connected in specific locations of the power grid. The underlying circuit is exercised by a number of random vector scenario in order to provide a distribution for the current drawn by each gate, and, therefore, by each individual current source attached to a specific power grid location. Each such current source set is processed through circuit-level power grid analysis to determine the current flowing through each interconnect element provided in the extracted power grid netlist, and, by using the layout data, calculate the current density J through each such element.

This process is repeated over all randomly generated current source sets in order to generate a current density distribution through each interconnect element. This distribution is then used to determine the average current density and the MTTF for the specific element through the use of Black's equation.

The results are then tabulated and sorted to provide the required feedback to the designers.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

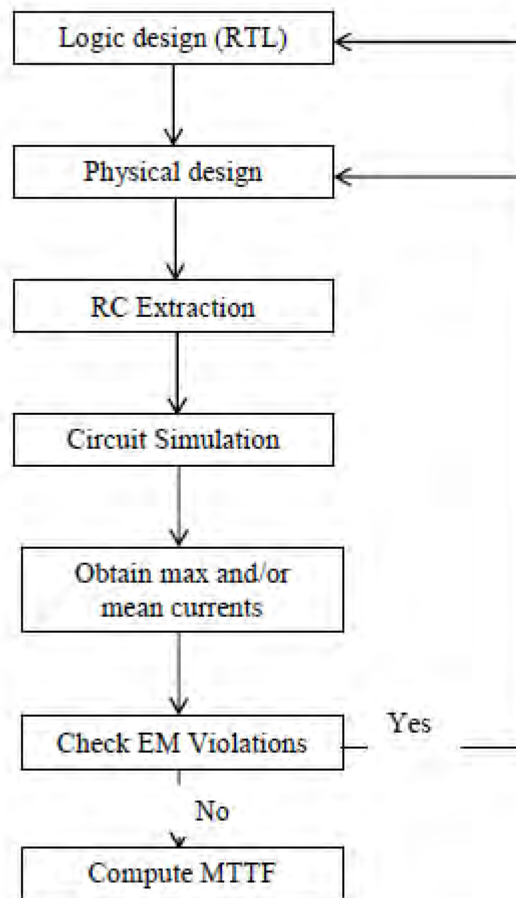


Figure VI Electromigration Workflow

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

4. Benchmarks and Results

4.1 DEF/LEF Introduction

A DEF (Design Exchange Format) file is an open specification for representing physical layout of an integrated circuit in an ASCII format. It contains the design-specific information of a circuit and is a representation of the design at any point during the layout process. DEF was developed by Cadence Design Systems. DEF conveys logical design data to, and physical design data from, place-and-route tools. Logical design data can include internal connectivity (represented by a netlist), grouping information, and physical constraints. Physical data includes placement locations and orientations, routing geometry data, and logical design changes for backannotation. They are usually generated by place and route tools. It is used in conjunction with the LEF File [10] [21].
A LEF (Library Exchange Format) file is an open specification for representing physical layout information on components of an integrated circuit in an ASCII format. The LEF contains library information for a class of designs. Library data includes layer, via, placement site type, macro cell definitions. The LEF file is strongly connected with the DEF file [10] [21].

4.2 Benchmarks

For the purpose of the master thesis two kind of benchmarks were used. The ISCAS benchmarks and some synthetic benchmark in order to prove the assumption that the CG method with the Fast Transform Preconditioner has great impact in time acceleration.
Firstly, the ISCAS '85 benchmark circuits are ten combinational networks provided to authors at the 1985 International Symposium on Circuits and Systems.1 They subsequently have been used by many researchers as a basis for comparing results in the area of test generation. Each circuit is characterized in the table below:

Circuit Name	Circuit Function	Total Gates	Input Lines	Output Lines	Faults
C432	Priority Decoder	160	36	7	524
C499	ECAT	202	41	32	758
C880	ALU and Control	383	60	26	942
C1355	ECAT	546	41	32	1574
C1908	ECAT	880	33	25	1879
C2670	ALU and Control	1193	233	140	2747
C3540	ALU and Control	1669	50	22	3428
C5315	ALU and Selector	2307	178	123	5350
C6288	16-bit Multiplier	2406	32	32	7744
C7552	ALU and Selector	3512	207	108	7550

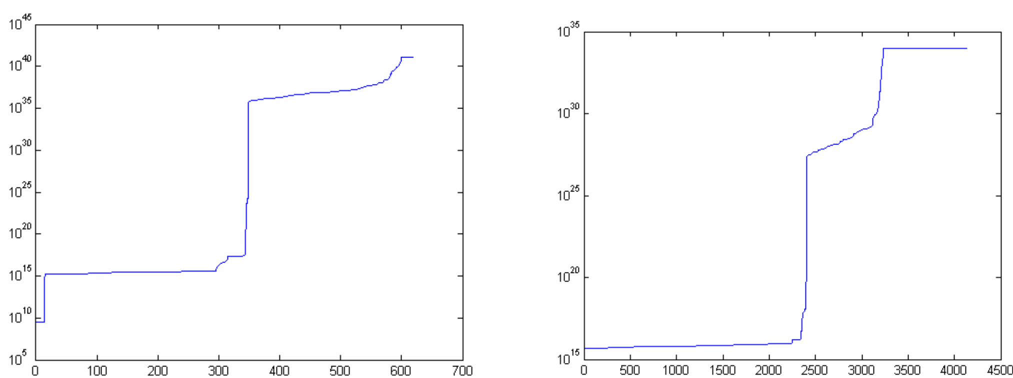
Table I ISCAS Benchmarks

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

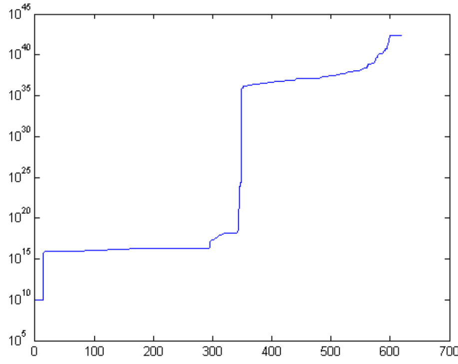
The ISCAS '85 netlist format was never formally documented; rather, it was distributed on magnetic tape along with a FORTRAN translator that would generate netlists in a few other formats. Now, a new translator written in C is being distributed with the ISCAS '85 benchmarks that outputs a single, generic-looking format which is more useful for translation to other formats (e.g., hilo, cadat). The ISCAS '85 format has, however, become viable despite its shortcomings. One reason for this is that it contains information not present in most other netlist formats. For instance the ISCAS '85 format lists each network node in levelized order, and this information may be lost when translating to other formats. Also, the ISCAS '85 format lists fanout branches separately as distinct nodes (with distinct names) and specifies the connectivity for each fanout branch. This is valuable for test generation purposes and must be extracted from other, more generic, netlists.es the connectivity for each fanout branch. This is valuable for test generation purposes and must be extracted from other, more generic, netlists.
Secondly, for the purpose of large scale full chip thermal analysis synthetic benchmarks were created with a MATLAB script, due to lack of benchmarks having millions of cells, in order to evaluate the proposed algorithm in large scale circuit.

4.3 Results

In order to evaluate the proposed algorithm we synthesized the power grid of the ISCAS85 benchmark circuits and applied random vectors at their inputs in order to create the required current excitations. We applied a temperature of 105ºC to all interconnect elements in the power grid. The results obtained were independent of the factor *A* since it is process dependent (but constant). The vectors applied to benchmark circuit c1355 had a 4GHz clock and its power supply network consisted of 618 elements. Benchmark circuit c6288 had a 100MHz clock and its power supply network consisted of 4133 elements.
After running the tool, the results were tabulated in logarithmic scale as the differences in MTTF for the power grid interconnect elements spanned over several orders of magnitude but none was less than a few years, which made all interconnect elements pass the required reliability threshold.



Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168



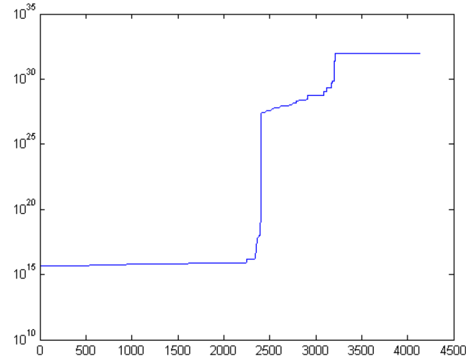


Figure 2. The x-axis depicts the number of the interconnects that had MTTF shown in the log y-axis of c1355

Figure 3. The x-axis depicts the number of the interconnects that had MTTF shown in the log y-axis of c6288

Figure 2 refers to the c1355 power grid and Figure 3 refers to c6288 power grid. In both figures the first plot illustrates the results for the maximum current case while the second shows the more realistic case by calculating the mean current. The results are presented in ascending order so that the leftmost result corresponds to the minimum MTTF value which is the worst case scenario.

Alongside with the ISCAS benchmarks, in order to evaluate the efficiency of the proposed methodology for full chip thermal simulation, we compared three methods for solving the linear systems for the thermal grid: : the PCG method with zero-fill Incomplete Cholesky preconditioner (ICCG), the proposed method of using PCG with the Fast Transform preconditioners (ET-FTCG), and CHOLMOD [6] which is a state-of-the-art direct solver for sparse SPD linear systems Each method was ported on a GPU platform and the only part that is executed on the CPU is the construction of the thermal grid preconditioner for ET-FTCG and ICCG. Subsequently, the CPU is responsible for transferring the appropriate data to the GPU. We have used the CUDA library [14] (version 4.2, along with CUBLAS, CUSPARSE and CUFFT libraries) for mapping the ICCG and the ET-FTCG algorithm on the GPU.
Due to the lack of a set of available benchmarks for full chip thermal analysis we have created a set of synthetic benchmarks, based on the theory that described in section 2 with size ranging from 153k to 10M nodes (Table II).

Benchmark	NumOfCuboids	Layers
Therm1	175	5
Therm2	320	5
Therm3	410	6
Therm4	500	7
Therm5	845	7
Therm6	946	7
Therm7	1118	8

Table II Benchmark Explanation

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

Where NumOfCuboids are the number of discretization in each axis, layers are the number of layers that the assumed benchmark has. Moreover, the matrix dimensions can be calculated by multiplying the NumOfCuboids in the square by the number of layers.
For the thermal grid, the length Δz of the grid rectangle was selected equal to the layer thickness (which can be variable), while the lengths Δx and Δy were chosen equal to the smallest routing width/pitch within a layer. All experiments were executed on a Linux workstation, comprising an Intel Core i7 processor running at 2.4GHz (6 cores and 24GB main memory) and an NVIDIA Tesla C2075 GPU with 5GB of main memory. Table III presents the results from the evaluation of the aforementioned methods on the set of benchmark circuits. The number Iter. is the number of iterations that the algorithm needed to converge while Time (s) refers to the time in seconds that needed to compute the final solution.
Comparing the iterative methods, we can observe that the proposed method was able to greatly reduce the number of iterations required for convergence. Compared with general purpose preconditioning methods such as Incomplete Cholesky factorization, the proposed preconditioners take into account the topology characteristics of the thermal grid. As a result, they are able to approximate them faithfully enough and reduce the required number of iterations. Moreover, owing to their inherent parallelism, the proposed preconditioners can utilize the vast amount of computational resources found in massively parallel architectures, such as GPUs. Thus, their efficacy is increased with the increasing circuit size. FTCG was able to achieve a speed-up ranging between 1.5X and 2.2X in CPU execution and 16X and 26.93X in GPU execution over ICCG.

Benchmark	ICCG		FTCG-CPU		FTCG-GPU	
	Iter.	Time (s)	Iter.	Time (s)	Iter.	Time (s)
Therm1	48	0.48	12	0.31	11	0.03
Therm2	57	1.98	15	1.23	12	0.08
Therm3	58	4.20	16	2.64	12	0.23
Therm4	67	8.35	17	4.51	12	0.31
Therm5	58	21.07	12	9.48	11	1.34
Therm6	59	28.27	16	17.94	11	1.60
Therm7	68	49.35	17	33.07	12	2.39

Table III Simulation Results

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

5. Conclusion and Future Work

Thermal and electromigration analysis s is among the most challenging problems facing the EDA industry today. The rapid reduction in the size of state of the art processors (current condition is beyond 20nm) is going to attract more attentions from the research community. Moreover, it will become necessary to modern reliability CAD tool to solve more complex problems and simultaneously offering more accurate result. To this end, we develop this techniques as an early research stage of the mentioned reliability tools.

Summarizing, the present work described two important reliability problems and proposed from our point of view accurate and fast computation models. Even though the present work is for the purpose of a master dissertation, all the implementation could be easily extend and compete several industrial tools. Furthermore, all the tools that were implemented could adapt to an industrial design flow.

Finally, in the future we plan to extend the current work toward the following directions:

- Microcooling thermal modeling: 3D-ICs bring about new challenges to chip thermal management due to their high heat densities. Micro-channel based liquid cooling and thermal through-silicon-vias (TSVs) have been adopted to alleviate the thermal issues in 3D-ICs. Thermal TSV (which have no electrical significance), enables higher interlayer thermal conductivity thereby achieving a more uniform thermal profile. While somewhat effective in reducing temperatures, they are limited by the nature of the heat sink. On the other hand, micro-channel based liquid cooling is significantly capable of addressing 3D IC cooling needs but consumes a lot of extra power for pumping coolant through channels. As a result it will be a good idea to study and extend the current work in the direction of taking into account designs with the presence of microcoolant flows.

- Extreme value theory engine: Probabilistic simulation of the electromigration is going to be a very computation expensive problem. In order to reduce the execution time of predicting MTTF, it will be very useful to apply an extreme value theory approach that reduces the computation time, but provides quite accurate results.

- Electromigration algorithm extensions: Along with that we plan to extend the algorithm in order to calculate accurately the factor A, taking into consideration the wire length and shape of the interconnection.

- Evaluation on large-scale actual industrial designs: All the benchmarks that have been mentioned in the previous sections was firstly the ISCAS benchmarks, that they are small designs and as a result those benchmarks do not have high on chip temperatures, and secondly the synthetic benchmarks that we created to demonstrate the speed up that can achieve the solution method that we adopt.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

- Integration of a graphic engine in the flow in order to visualize the results and offering capabilities like an industrial tool.

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

6. References

[1] Alam, S. M., Troxel, D. E., & Thompson, C. V. (2003). Layout-specific Circuit Evaluation in Three-dimensional Integrated Circuits. *Analog Integrated Circuits and Signal Processing, An International Journal. Kluwer Academic Publishers*, 199- 206.

[2] Banerjee, K., Mehrotra, A., Sangiovanni-Vincentelli, A., & and C. Hu. (1999). On thermal effects in deep sub-micron vlsi interconnects. *Design Automation Conference*.

[3] Black, J. (1969). Electromigration —A brief survey and some recent results. *IEEE Trans. on Electron Devices 16 (4)*, 338–347.

[4] Black, J. (1969). Electromigration Failure Modes in Aluminium Metallization for Semiconductor Devices. *Proc. of the IEEE 57 (9)*, 1587–94.

[5] Borovkov, A. A. (1967). On limit laws for service processes in multi-channel systems. *Siberian Mathematics Journal*, 746-763.

[6] Chen, Y., Davis, T. A., Hager, W. W., & Rajamanickam, S. (2008). Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Up-date/Downdate. *ACM Trans. Math. Softw., vol. 35, no. 3*, 22:1-22:14.

[7] Cheng, Y.-K., Raha, P., Teng, C.-C., Rosenbaum, E., & Kang, S.-M. (1998). \Illiads-t: an electrothermal timing simulator for temperature-sensitive reliability diagnosis of cmos vlsi chips. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, vol. 17, no. 8*, 668-681.

[8] Cherry, Y., Hau-Riege, S., Alam, S. M., Thompson, C. V., & Troxel, D. (1999). ERNI: A Tool for Technology-Generic Circuit-Level Reliability Projections. *Interconnect Focus Center Annual Review*.

[9] Hu, C. (1992). (Berkeley Reliability Simulator) BERT: an IC reliability simulator. *Microelectronics Journal, 23*, 97-102.

[10] LEF/DEF Language Reference. (n.d.).

[11] Li, P., Pileggi, L. T., Asheghi, M., & Chandra, R. (2006). IC Thermal Simulation and Modeling via Efficient. *IEEE transactions on computer-aided design of integrated circuits and systems, vol. 25, no. 9*, 1763-1776.

[12] Lit, P., Pileggi, L. T., Asheghi, M., & Chandra, R. (n.d.). Efficient Full-Chip Thermal Modeling and Analysis. 2004.

[13] Intel Math Kernel Library. [Online]. Available: http://software.intel. com/en-us/articles/intel-mkl/ . (n.d.).

[14] NVIDIA CUDA Programming Guide, CUSPARSE, CUBLAS, and CUFFT Library User Guides. [Online]. Available: http://developer.nvidia.com/nvidia-gpu-computing- documentation . (n.d.).

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168

[15] Ozisik, M. N. (n.d.). Heat transfer: A basic approach. *NY:McGraw-Hill.*

[16] Stamoulis, G. I. (1996). A Monte-Carlo approach for the accurate and efficient estimation of average transition probabilities in sequential logic circuits. *Proceedings of the 1996 Custom Integrated Circuits Conference*, 221-224.

[17] Teng, C.-C., Cheng, Y., Rosenbaum, E., & Kang, a. S. (1997). iTEM: A Temperature-Dependent Electromigration Reliability Diagnosis Tool. *IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems, vol. 16*, 882-893.

[18] Ting, L., May, J., Hunter, W., & McPherson, J. (1993). AC electromigration characterization and modeling of multilayered interconnects. *Proceedings of the IEEE International Reliability Physics Symposium*, 311-316.

[19] Wang, B., & Mazumder, P. (325-344). Accelerated Chip-Level Thermal Analysis Using. *IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, VOL. 26, NO. 2*, 2007.

[20] Wang, T.-Y., & Chen, C. C.-P. (2002). 3-D Thermal-ADI: A Linear-Time Chip Level. *IEEE transactions on computer-aided design of integrated circuits and systems, vol. 21, no. 12*, 1434-1445.

[21] Wikipedia, Design Exchange Format, Library Exchange Format. (n.d.).

[22] Yang, Y., Gu, Z. (., Zhu, C., Dick, R. P., & Li Shang. (86-99). ISAC: Integrated Space-and-Time-Adaptive. *IEEE transactions on computer-aided design of integrated circuits and systems, vol. 26, no. 1*, 2007.

[23] Zhan, Y., & Sapatnekar, S. S. (1661-1775). High-Efficiency Green Function-Based Thermal Simulation Algorithms. IEEE *transactions on computer-aided design of integrated circuits and systems, vol. 26, no. 9*, 2007.

[24] Daloukas, K. Evmorfopoulos, N. , Drasidis, G., Tsiampas M., Tsompanopoulou, P., Stamoulis, G.I.(2012) Fast Transform-based preconditioners for large-scale power grid analysis on massively parallel architectures. 2012 IEEE/ACM International Conference on Computer-Aided Design (ICCAD),384-961

Institutional Repository - Library & Information Centre - University of Thessaly
23/09/2024 09:17:33 EEST - 18.118.186.168