3-D Floorplanning Algorithm to Minimize Thermal Interactions

Boris Vaisband, *Student Member, IEEE*, and Eby G. Friedman, *Fellow, IEEE*

Department of Electrical and Computer Engineering,
University of Rochester, Rochester, New York 14627 USA.
[bvaisban,friedman]@ece.rochester.edu.

***Abstract*— An algorithm for including relative thermal interactions among different circuit modules within a 3-D system is introduced in this paper. Application of the proposed algorithm on MCNC and GSRC benchmark circuits is presented. The thermal behavior of a heterogeneous 3-D structure, consisting of a different number of modules and substrate materials, is evaluated to emulate the heat transfer characteristics of practical heterogeneous 3-D systems. The algorithm lowers thermal interactions between different modules while maintaining the peak temperature within a practical range. The thermal characteristics of the floorplan are evaluated using HotSpot and HotSpot Detailed 3-D and compared to a random floorplan. The recorded peak temperatures are within the practical range of on-chip temperatures.**

I. Introduction

The integrated circuits community is driven to develop heterogeneous applications. Different types of circuits (*e.g.*, sensors, analog, digital, and memory) are integrated in these applications, producing a complex environment of crosstalk and global signaling [1]. A three-dimensional (3-D) structure is an effective platform for integrating these diverse circuits [2], [3]. Different layers, optimized for specific circuits, can be independently manufactured and stacked to create a 3-D heterogeneous system. Unique requirements can be addressed on a per layer basis (*e.g.*, different substrates, individually optimized power networks, and local synchronization). An additional benefit of 3-D integration is the short vertical dimension that can alleviate global signaling issues within 2-D circuits while dissipating less power. These long global lines are replaced by short (*i.e.*, 20 to 80 μm [4]) vertical connections which exhibit significantly lower impedance.

Efficiently moving the heat generated within a circuit to the heat sink is an important objective in heterogeneous 3-D ICs where different layers are stacked on top of one another and heat is trapped within the 3-D structure. Thermal congestion within integrated circuits has been researched extensively, but remains a topic of concern in deeply scaled integrated circuits. The electrical characteristics of on-chip devices degrade with increasing temperature due to lower mobility carriers [5] and threshold voltage drift in transistors [6], as shown in Figure 1. The increase in temperature leads to significant degradation in performance and possible functional failure.

This research is supported in part by the Binational Science Foundation under Grant No. 2012139, the National Science Foundation under Grant No. CCF-1329374, the IARPA under Grant No. W911NF-14-C-0089, and by grants from Qualcomm, Cisco Systems, and Intel.

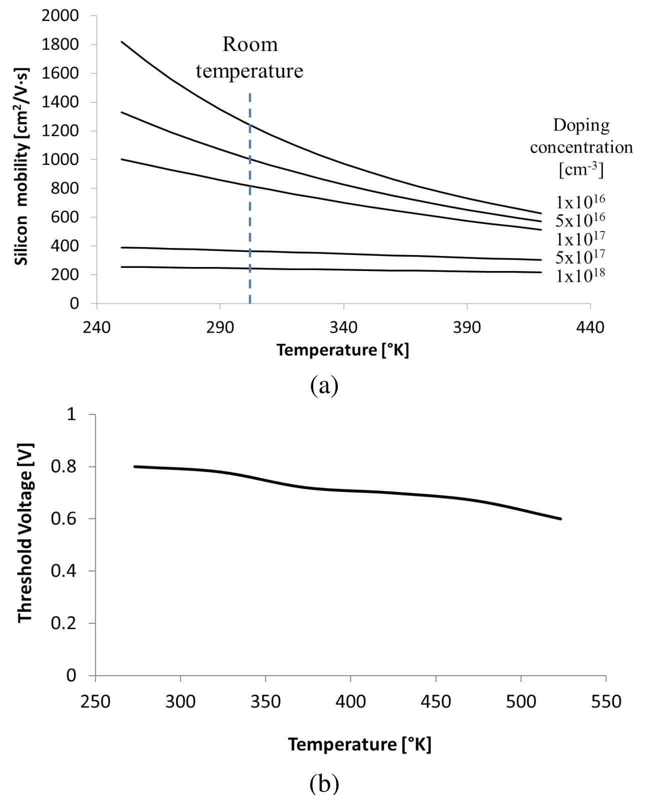

Fig. 1. Effect of increasing temperature on (a) mobility [5], and (b) threshold voltage [6].

Different solutions have been proposed to mitigate thermal congestion issues in 3-D ICs (*e.g.*, thermal TSVs, thermoelectric devices [7], and liquid cooling [8]). At the algorithmic level, thermal aware 3-D floorplanning algorithms have recently been developed [9], [10] to lower the peak temperature within a system. However, lowering the maximum temperature does not guarantee correct functionality and high performance because temperature is the measure of heat at a specific location and does not describe the effect of this heat on other locations.

As exemplified in Figure 2, the maximum temperature, denoted as T_{max}, is generated in module A, while a temperature denoted as T_b ($T_b < T_{max}$) is generated in module B which

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

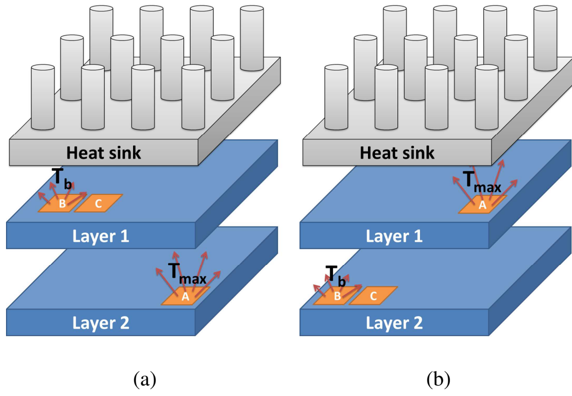

Fig. 2. Placement of three modules, A, B, and C, (a) initial state, and (b) after applying a temperature driven floorplan algorithm. T_{max} and T_b are the temperatures generated, respectively, in modules A and B.

is in close proximity to module C. Module C contains a temperature sensitive element which is greatly affected by the heat generated in module B. The *thermal interaction between modules* B and C is not considered by thermal aware algorithms that aim to lower the peak temperature. The result is functional failure or degradation in performance of the circuit in module C. Note that the thermal interaction between the different modules is the key issue. The focus of the work presented in this paper is a 3-D floorplanning algorithm that minimizes these thermal interactions.

The rest of the paper is composed of the following sections. Thermal interaction as the cost function within the proposed algorithm is presented in Section II. The proposed algorithm is reviewed in Section III. Evaluation of the algorithm on benchmark circuits is presented in Section IV. Some conclusions are offered in Section V.

II. Thermal Interactions

Each layer of a 3-D structure is typically divided into modules. To design circuits in a thermally efficient manner, each module is assigned two parameters: thermal aggressiveness t_a and thermal sensitivity t_s. The thermal aggressiveness is based on the heat flux generated within each module as well as the material properties of the module and surrounding area. This parameter characterizes the thermal effect of a module on neighboring modules. A thermal sensitivity parameter describes the tolerance to heat of a specific module. Issues such as delay uncertainty, carrier mobility, noise margin, threshold voltage variations, and dynamic range can affect the thermal sensitivity of a module. This parameter characterizes the sensitivity of a module to thermal coupling from neighboring modules. The parameters t_a and t_s range from 0 to 1 and are integrated within a floorplanning algorithm to determine the most thermally efficient location of each module. The thermal influence between modules i and j is defined in (1) as

$$T_{inf}^{i,j} \equiv t_a^i \cdot t_s^j + t_a^j \cdot t_s^i \ . \tag{1}$$

The thermal influence ranges from 0 to 2, where the lower bound considers those modules that do not thermally influence another module (and may be placed in close proximity to each other), and the upper bound considers those modules that exhibit a significant influence on another module (and must be placed far from each other). Note that thermal influence is not a vector, therefore $T_{inf}^{i,j} = T_{inf}^{j,i}$. Each module can be both thermally aggressive and thermally sensitive depending upon the circuit and the electrical and thermal characteristics. The thermal influence therefore describes the mutual influence between modules, and is expressed as the sum of the individual effects in (1).

Another important aspect of thermal interaction is the characteristics of the thermal path between modules i and j, described by the thermal resistance $R_{th}^{i,j}$ $[°C/W]$ [3], [11],

$$R_{th}^{i,j} = \frac{1}{k} \cdot \frac{\Delta x}{A} \ . \tag{2}$$

The thermal conductivity k $[W/m°C]$ is a property of the material, Δx $[\mu m]$ is the length of the thermal path between two modules, and A $[m^2]$ is the surface area of the thermal path through which the heat is transferred. Similar to thermal influence, the thermal resistance is not a vector, therefore $R_{th}^{i,j} = R_{th}^{j,i}$. The thermal interaction between modules i and j is defined in (3) as

$$T_{int}^{i,j} \equiv \frac{T_{inf}^{i,j}}{R_{th}^{i,j}} \ . \tag{3}$$

Thus, $T_{int}^{i,j} = T_{int}^{j,i}$ as the result of the division of two scalars. The thermal interaction $[W/°C]$ is similar to thermal conduction adjusted according to the thermal influence between modules. Substituting (1) into (3),

$$T_{int}^{i,j} = \frac{t_a^i \cdot t_s^j + t_a^j \cdot t_s^i}{R_{th}^{i,j}} \ . \tag{4}$$

The cost function of the proposed algorithm is to minimize $max(T_{int})$ and therefore minimize the conduction of heat between the thermally aggressive and thermally sensitive modules. This algorithm does not necessarily lower the maximum temperature, but generates a 3-D floorplan less susceptible to functional failure and performance degradation.

III. Proposed Algorithm

The proposed thermal interaction driven 3-D floorplan algorithm generates a thermally efficient placement of all component modules making up a 3-D system. The circuits on each layer are divided into modules and placed on a layer within a 3-D system. The layer is divided into regions. A single module can be placed within each region. A thermal path exists from any region i to all other regions on all layers. The thermal path between any two regions is characterized by the thermal resistance between those regions. For the inter-layer thermal paths, the thermal resistance of the TSVs is included in the characterization of the thermal path. The number of modules k is user defined according to the number of functional modules within a circuit. The number of regions s is based on the layer area and number of modules ($s > k$). White spaces are incorporated and reserved for TSV placement. After the

TABLE I

APPLICATION OF PROPOSED ALGORITHM ON MCNC AND GSRC BENCHMARK CIRCUITS.

Benchmark	Number of layers	Modules per layer	Die area $[mm^2]$	T_{int}^{wc} $[W/^{\circ}C]$	T_{int}^{max} $[W/^{\circ}C]$	η	Run time $[s]$
ami33	2	33	41.5	$5.71 \cdot 10^{-5}$	$1.32 \cdot 10^{-4}$	0.43	0.4
	3			$5.58 \cdot 10^{-5}$	$1.73 \cdot 10^{-4}$	0.32	0.8
	5			$1.07 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$	0.43	2.9
ami49	2	49	1539.4	$1.91 \cdot 10^{-5}$	$1.27 \cdot 10^{-4}$	0.15	1.3
	3			$2.21 \cdot 10^{-5}$	$8.77 \cdot 10^{-5}$	0.18	3.1
	5			$2.36 \cdot 10^{-5}$	$1.24 \cdot 10^{-4}$	0.19	13.8
n100	2	100	6.6	$5.56 \cdot 10^{-5}$	$2.15 \cdot 10^{-4}$	0.26	9.6
	3			$4.68 \cdot 10^{-5}$	$1.82 \cdot 10^{-4}$	0.26	37
	5			$5.66 \cdot 10^{-5}$	$1.84 \cdot 10^{-4}$	0.31	240
n200	2	200	6.2	$4.17 \cdot 10^{-5}$	$1.47 \cdot 10^{-4}$	0.28	121
	3			$4.74 \cdot 10^{-5}$	$1.94 \cdot 10^{-4}$	0.24	594
	5			$5 \cdot 10^{-5}$	$1.77 \cdot 10^{-4}$	0.28	4,626
n300	2	300	9.3	$5.2 \cdot 10^{-5}$	$1.77 \cdot 10^{-4}$	0.29	637
	3			$5.12 \cdot 10^{-5}$	$2.15 \cdot 10^{-4}$	0.24	3,055
	5			$5.3 \cdot 10^{-5}$	$1.94 \cdot 10^{-4}$	0.27	20,884

number of modules is determined, a thermal influence array of length $k \cdot (k-1)/2$ is produced according to (1) and sorted in descending order. A thermal resistance array of length $s \cdot (s-1)/2$ is also produced based on the thermal conductivity of the material, surface area through which the heat is transferred, and distance between regions. These arrays are used within the algorithm to compute the thermal interactions between any two modules. A matrix representation of this system is

$$\begin{pmatrix} T_{int}^{1,1} & T_{int}^{1,2} & \cdots & T_{int}^{1,\frac{s\cdot(s-1)}{2}} \\ T_{int}^{2,1} & T_{int}^{2,2} & \cdots & T_{int}^{2,\frac{s\cdot(s-1)}{2}} \\ \vdots & \vdots & \ddots & \vdots \\ T_{int}^{\frac{k\cdot(k-1)}{2},1} & T_{int}^{\frac{k\cdot(k-1)}{2},2} & \cdots & T_{int}^{\frac{k\cdot(k-1)}{2},\frac{s\cdot(s-1)}{2}} \end{pmatrix} = \begin{pmatrix} T_{inf}^{1,1} \\ T_{inf}^{2,1} \\ \vdots \\ T_{inf}^{\frac{k\cdot(k-1)}{2},1} \end{pmatrix} \times \begin{pmatrix} R_{th,rec}^{1,1} & R_{th,rec}^{1,2} & \cdots & R_{th,rec}^{1,\frac{s\cdot(s-1)}{2}} \end{pmatrix}. \quad (5)$$

The matrix T_{int} of size $[\frac{k\cdot(k-1)}{2} \times \frac{s\cdot(s-1)}{2}]$ represents all possible thermal interactions within a system, T_{inf} is a column vector of size $[\frac{k\cdot(k-1)}{2} \times 1]$ that represents the thermal influence among all modules, and $R_{th,rec}$ is a row vector of size $[1 \times \frac{s\cdot(s-1)}{2}]$ that represents the reciprocal of the thermal resistances among all regions on all layers. A module i may only be placed on the layer assigned to module i; therefore, only the free space on that specific layer is available for module i. During each iteration of the algorithm, the pair of unplaced modules with the greatest thermal influence is placed. The regions are determined by an exhaustive search (complexity $O(s^2)$) to minimize the thermal interaction between the modules placed in the current iteration with all of the previously placed modules (on all layers). Hence, each module is placed in a region least sensitive to thermal interactions with all of the other modules. The output of the algorithm is a legal floorplan, where each module is placed on the assigned layer. After the placement iterations are completed, the worst case thermal interaction (T_{int}^{wc}) of the resulting floorplan is determined.

IV. EVALUATION OF ALGORITHM ON BENCHMARK CIRCUITS

The algorithm has been tested on existing MCNC and GSRC benchmark circuits to ensure correct functionality and evaluate systems with many modules. The evaluations are performed on an Intel Core i5-2410M CPU @ 2.30 GHz machine with Windows 7 64-bit OS and the algorithm has been implemented in MATLAB R2011b. The thermal resistance is determined according to the thermal conductivity of silicon at $100^{\circ}C$ ($k = 113\ W/m^{\circ}C$), size of the modules, and distance between the different placement locations within a layer. The thermal aggressiveness and thermal sensitivity in the benchmark circuits are randomly generated. To assess the quality of the resulting floorplan, a quality figure of merit η is used,

$$\eta \equiv \frac{T_{int}^{wc}}{T_{int}^{max}} \ , \quad (6)$$

where T_{int}^{max} is the maximum thermal interaction defined as the greatest thermal influence over the smallest thermal resistance. A smaller value of η indicates a better result.

A. MCNC and GSRC benchmark circuits

MCNC and GSRC benchmark circuits are evaluated for different number of layers, and the quality parameter η from (6) and run time are recorded. The results are listed in Table I. Each benchmark circuit is simulated for two, three, and five layers with a similar number of modules on each layer. For each test case, white spaces are inserted within each layer to accommodate the electrical and thermal TSVs carrying, respectively, the electrical and thermal signals. To emphasize the applicability of the algorithm to heterogeneous systems, evaluations are presented in subsection IV-B for different substrate materials and number of modules on each layer.

TABLE II

PARAMETERS OF HETEROGENEOUS 3-D SYSTEM TEST CASE.

Layer	Modules on layer	Substrate material	White spaces
1	49	Si	72
2	33	GaAs	88
3	100	Ge	21

B. Heterogeneous 3-D systems

The 3-D structure is favorable for heterogeneous applications consisting of different types of circuits on different layers. An additional test case is considered to emulate a heterogeneous 3-D system with different types of circuits on each layer. The proposed algorithm is applicable to heterogeneous systems by choosing different thermal parameters for each layer. These parameters include the thermal substrate resistance and number of modules for each layer. The test case parameters are listed in Table II. The system consists of three layers with a die area of 25 mm^2. Different substrate materials are assigned to each layer to emulate different circuit types on each layer (*e.g.*, Si for digital CMOS, GaAs for RF circuits, and Ge for photovoltaic cells). The thermal properties (*i.e.*, t_a and t_s) are randomly generated.

The algorithm returns a legal floorplan for each layer, where the placement is chosen to minimize the thermal interaction of every module with all other modules (on all layers). The worst case thermal interaction of the heterogeneous test case is $5.8 \cdot 10^{-5}$ $W/°C$ and $\eta = 0.25$. The run time of the algorithm is 14 s.

C. HotSpot simulations

To ensure that the peak temperature of the 3-D floorplans generated by the proposed algorithm is within the operational range, the circuits are simulated by HotSpot [12] and HotSpot Detailed 3-D [13] tools to determine the peak temperature.

A four layer 3-D structure evaluated on the ami33 benchmark circuit for each layer has been generated by the proposed algorithm. The generated floorplan is used as an input for HotSpot Detailed 3-D [13]. Similar to [9] and [10], the power density ranges from 10^5 to 10^7 W/m^2 and is randomly generated for each module. The white spaces are modeled with different thermal resistances to emulate that 10% of this area is occupied by TSVs. A random floorplan is initially evaluated, and a peak temperature of $139°C$ is recorded. A maximum temperature of $136.8°C$ is noted for the generated floorplan. Identical power densities are used for both floorplans. Both peak temperatures are within the range of practical temperatures for 3-D ICs [9], [10]. The worst case thermal interaction for the floorplan generated by the proposed algorithm is $9.89 \cdot 10^{-5}$ $W/°C$ and $\eta = 0.39$.

V. CONCLUSIONS

The issue of thermal congestion within 3-D structures is the focus of this paper. Thermal congestion within 3-D ICs can lead to degradation in performance and circuit failure. Not all circuits however are sensitive to thermal effects and not all circuits generate significant amounts of heat. To ensure high performance and correct functionality, the thermal interaction between modules should be considered. Algorithms that minimize the maximum system temperature [9], [10] do not necessarily achieve these goals. The primary innovation of the proposed algorithm is to lower thermal interactions between the different thermally aggressive and thermally sensitive parts of a 3-D IC.

A thermal interaction floorplanning algorithm for 3-D heterogeneous systems is described in this paper and an evaluation of this algorithm on several benchmark circuits is provided. The algorithm places the modules on each layer according to the thermal interactions exhibited by the surrounding modules. As listed in Table I, significant improvement in the worst case thermal interaction is possible ($\eta < 0.5$ for all test cases).

The HotSpot [12] and HotSpot Detailed 3-D [13] tools are used to ensure that the peak temperature of the floorplans produced by the proposed algorithm is below a practical limit. The peak temperature in the generated floorplan is $136.8°C$, within the range of peak temperatures recorded in previous work [9], [10]. This peak temperature is lower than the peak temperature recorded for a randomly generated floorplan.

REFERENCES

[1] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*, McGraw-Hill Publishers, 2012.

[2] V. F. Pavlidis and E. G. Friedman, *Three-Dimensional Integrated Circuit Design*, Morgan Kaufmann, 2009.

[3] B. Vaisband, I. Savidis, and E. G. Friedman, "Thermal Conduction Path Analysis in 3-D ICs," *Proceedings of the IEEE Symposium on Circuits and Systems*, pp. 594 – 597, June 2014.

[4] ITRS Technology Working Groups, "International Technology Roadmap for Semiconductors (ITRS)," 2011.

[5] N. D. Arora, J. R. Hauser, and D. J. Roulston, "Electron and Hole Mobilities in Silicon as a Function of Concentration and Temperature," *IEEE Transactions on Electron Devices*, Vol. 29, No. 2, pp. 292 – 295, February 1982.

[6] J. J. Tzou, C. C. Yao, R. Cheung, and H. Chan, "The Temperature Dependence of Threshold Voltages in Submicrometer CMOS," *IEEE Electron Device Letters*, Vol. 6, No. 5, pp. 250 – 252, May 1985.

[7] H. Chen, L. Hsu, and X. Wei, "A Novel VLSI Technology to Manufacture High-Density Thermoelectric Cooling Devices," *Proceedings of the International Workshop on THERMal INvestigations of ICs and Systems*, pp. 1 – 6, January 2008.

[8] Z. Li and M. S. Bakir, "Electrical and Fluidic Microbumps and Interconnects for 3D-IC and Silicon Interposer," *Proceedings of the IEEE SOC Conference*, pp. 159 – 164, Spetember 2012.

[9] Z. Pingqiang, M. Yuchun, L. Zhouyuan, R. P. Dick, S. Li, Z. Hai, H. Xianlong, and Z. Qiang, "3D-STAF: Scalable Temperature and Leakage Aware Floorplanning for Three-Dimensional Integrated Circuits," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 590 – 597, November 2007.

[10] J. Cong, W. Jie, and Z. Yan, "A Thermal-Driven Floorplanning Algorithm for 3D ICs," *Proceedings of the IEEE/ACM International Conference on Computer Aided Design*, pp. 306 – 313, November 2004.

[11] W. Huang, *HotSpot - A Chip and Package Compact Thermal Modeling Methodology for VLSI Design*, Ph.D. Thesis, University of Virginia, January 2007.

[12] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-Aware Microarchitecture," *Proceedings of the IEEE International Symposium on Computer Architecture*, pp. 2 – 13, May 2003.

[13] M. Jie, K. Kawakami, and A. K. Coskun, "Optimizing Energy Efficiency of 3-D Multicore Systems with Stacked DRAM under Power and Thermal Constraints," *Proceedings of the ACM/IEEE Design Automation Conference*, pp. 648 – 655, March 2012.