Coupled Simulation of Device Performance and Heating of Vertically Stacked Three-Dimensional Integrated Circuits

Akin Akturk, Neil Goldsman and George Metze†
Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA
†Laboratory of Physical Sciences, College Park, MD 20742, USA
Email: {akturka,neil,metze}@glue.umd.edu

***Abstract* - We have developed a method for calculating the temperature distribution of a three-dimensional (3D) integrated circuit (IC) and the performance of a single device self-consistently. At the device level, we resolve effects of channel temperatures and thermal boundary conditions on device performance. Thus we obtain non-isothermal device characteristics for a representative device of the technology node used to fabricate the 3D-IC. At the 3D-IC level, we first approximate the average heat generation of each device on the chip using a statistical Monte Carlo method. We second determine effects of the chip layout and the fabrication materials on thermal coupling. Next we calculate thermal profile of the 3D-IC in conjunction with the individual device operations, 3D-IC layout and full-chip workload statistics. Our technique offers a numerical method to isolate affects of chip layout, floor-plan and operational activity on 3D-IC thermal profile. Thus it enables designers to pinpoint potential hotspots, and test new design paradigms for cooling methods.**

I. INTRODUCTION

As interconnect delays limit the speed of today's electronic systems, researchers are merging systems on a chip to overcome speed and real estate problems [1]. However, the device and full-chip heating is emerging as a primary limiting factor to achieving performance goals, or even functionality in 3D-ICs. Heating, which is one of the biggest problems even for planar chips [2-3], is aggravated within the layers of 3D-ICs due to restricted heat flow through the insulation layers. Thus, we propose a methodology for numerical thermal analysis of vertically stacked 3D-ICs, as the one shown in Fig. 1. We obtain the thermal profile of 3D-ICs in accordance with their individual device operations. The proposed method enables calculation of the thermal profile of the 3D-IC with a desired resolution that can be as small as a single MOSFET. To obtain accumulated effects of device operation on the heating of the entire 3D-IC, we employ a quantum device simulator [3] along with a solver for a KCL-type thermal network. In addition, to include operational variations due to floor plans and software applications for each 3D-IC's device, we employ a Monte Carlo-type algorithm. Our calculated results indicate that layers that form a vertically stacked 3D-IC, especially the ones in the middle, may severely be affected adversely, with temperatures reaching more than one hundred degrees Celsius above the ambient.

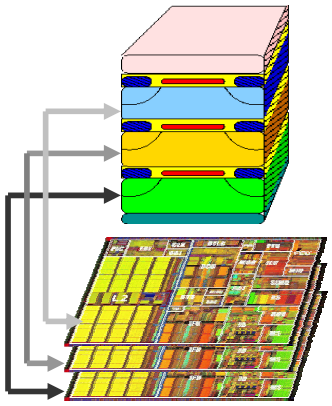

Figure 1: 3D-IC: Each of the three layers is a chip modeled after a Pentium III [4].

II. THERMAL 3D-IC HEATING MODEL

To determine temperature profile of complex 3D-ICs, we develop a lumped thermal network model, which calculates the heat flow between devices on the 3D-IC, and the distribution of the heat generated.

A. Lumped Model for Heat Flow

To model the heat flow between devices, we start from the differential heat flow equation written below:

$$C\frac{\partial T}{\partial t} = \nabla.\left(\kappa \nabla T\right) + H \qquad (1)$$

Using Eqn. (1), spatial temperature, T, distribution can be found by solving the steady-state equation, which is a function of heat flux ($-\kappa\nabla T$, where κ is the thermal diffusion constant.) and heat generated, H. In addition, thermal transients can be included using the time derivative of temperature weighted by heat capacity, C.

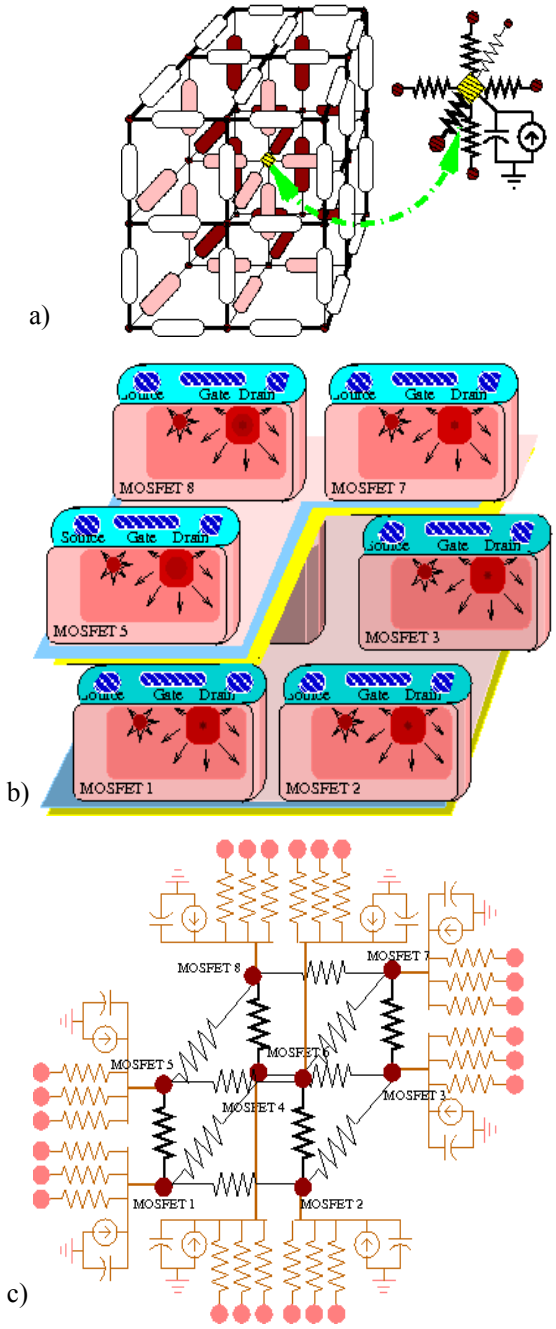

Figure 2: a) Lumped thermal network for a 3D-IC. Each thermal node is connected to six other nodes and ground via thermal resistors, and heating source and capacitance, respectively. b) A close-up look at two layers of a 3D-IC with eight devices (The closest MOSFET (6) to view is not plotted. The furthest MOSFET (4) is partially shown.) Heat generated peaks in the channel near the source and drain junctions. c) Thermal network between the devices in b). Each device is also connected to three other neighboring devices, and to ground.

We develop a lumped heat flow model using Eqn. (1), since the application of the differential heat flow equation (1) renders the solution for the 3D-IC intractable due to our goal of having spatial resolution of a single device, and the 3D-IC simultaneously.

To develop our lumped thermal model, we cover a device on the 3D-IC by a rectangular prism. We then assume each rectangular prism, equivalently a MOSFET in Fig. 2(b), can be associated with a point at its channel that has a single temperature value attributed to it. (Our calculations indicate that heat generation peaks in the channel near the source and drain junctions, with a global maximum near the drain junction. Thus one can assign the center of thermal activity to a point in the device channel near the drain junction. Additionally, for a bulk MOSFET peak channel lattice temperatures are less than one percent higher than those of the terminals enabling the use of the concept of center of thermal activity.)

We then apply Kirchoff's transformation to Eqn. (1) to move the thermal diffusion constant outside the gradient term. (Kirchoff's transformation is not necessary for obtaining a lumped heat flow model. However, it facilitates the solution by explicitly putting the burden of temperature dependency of thermal diffusion constant on the transformed temperature [5].) Below, we write Kirchoff's transformation, which allows us to utilize a constant value for the $\kappa(T_o)$ in Eqn. 1.

$$\overline{T} = T_o + \frac{1}{\kappa(T_o)} \int_{T_o}^{T} \kappa(\tau) d\tau \qquad (2)$$

We next integrate the heat flow equation around a rectangular prism mentioned before. Evaluations of the integrals yield KCL-type nodal equations, which for each node equate the net resistive heat flow between six neighboring nodes, and capacitive flow to the ambient to the heat generated at that node by the corresponding device (i,j,k). Therefore we can describe the system by thermal resistances, capacitances and sources using an electrical analogy. We depict the resulting thermal network in Fig. 2, and write down the corresponding KCL equations that describe the system below:

$$C^{th}_{i,j,k} \frac{(\overline{T}^{l}_{i,j,k} - \overline{T}^{l-1}_{i,j,k})}{\Delta t} + \frac{(\overline{T}^{l}_{i,j,k} - \overline{T}^{l}_{i\pm 1,j,k})}{R^{th}_{i\pm \frac{1}{2},j,k}} + \frac{(\overline{T}^{l}_{i,j,k} - \overline{T}^{l}_{i,j\pm 1,k})}{R^{th}_{i,j\pm \frac{1}{2},k}} + \frac{(\overline{T}^{l}_{i,j,k} - \overline{T}^{l}_{i,j,k\pm 1})}{R^{th}_{i,j,k\pm \frac{1}{2}}} = I^{l}_{i,j,k}(T^{l-1}_{i,j,k}) \qquad (3)$$

Using electrical analogy, $\overline{T}$, R^{th}, C^{th} and I are analogous to voltage, resistance, capacitance and current source, respectively. Moreover, thermal resistance can be calculated using physical lengths, 3D-IC layout and taking thermal diffusion constant as conductivity.

B. Distribution Model for Heat Generated

From 3D-IC layout, we can determine values for the thermal resistances and capacitances. To solve Eqn. (3), we also need to calculate values for heat generated by each device. However, there are millions of devices on a 3D-IC, with operating conditions differ for each of them depending on the software application and chip floor plan. Thus we use a Monte Carlo-type methodology to extend our calculated device results to the 3D-IC.

To determine effects of the 3D-IC floor plan, we group devices into a few functional blocks such as clock, cache, etc. We then find the percentage of the power consumed by each block. Next, we calculate the activity level of a device in a block relative to devices in other blocks by finding the percentage power densities in each block. We then associate these percentage power per areas with a probability density function. Using this density function, we obtain values for the percentage of the total steady-state power used by a device in that block. Therefore we determine the percentage workload of each unit and device.

C. Performance Model for MOSFET

$$\nabla^2\phi = -\frac{q}{\varepsilon}\left(p - n + D\right) \tag{4}$$

$$\frac{\partial n}{\partial t} = \nabla.\left(-n\mu_n\nabla\phi + \mu_n V_{TH}\nabla n\right) + GR_n \tag{5}$$

$$\frac{\partial p}{\partial t} = \nabla.\left(p\mu_p\nabla\phi + \mu_p V_{TH}\nabla p\right) + GR_p \tag{6}$$

$$C\frac{\partial T}{\partial t} = \nabla.\left(\kappa\nabla T\right) + H \tag{7}$$

$$E_i\psi_i(y) = -\frac{\hbar^2}{2m^*}\frac{d^2\psi_i(y)}{dy^2} - q\phi(x,y)\psi_i(y) \tag{8}$$

$$n = \frac{m^* kT}{\pi\hbar^2}\sum_i \left|\psi_i\right|^2 \ln\left(1 + e^{(E_F - E_i)/kT}\right) \tag{9}$$

We use a quantum device solver [3] to obtain device performance details including current-voltage characteristics and heat generated at different temperatures. Our device simulator solves quantum corrected semiconductor equations (Poisson equation 4 followed by the continuity equations for electrons 5 and holes 6, and the heat flow equation 7), which are coupled to the Schrödinger equation 8 and electron density equation 9 [3]. In addition, we resolve non-isothermal device behavior by explicitly changing the values of mobility, saturation velocity, built-in potentials, thermal voltage, intrinsic carrier concentration, bandgap and thermal diffusion constant. Using our simulator, we calculate steady state heat generated (Joule heating) as a function of temperature. We then input these results into the left-hand side of Eqn. (3).

D. Coupled Model for 3D-IC Heating

We first calculate values for thermal resistances and capacitances using the 3D-IC layout. We then determine workload statistics of each 3D-IC's device and unit. We decide on a distribution model for heat generated to approximate the percentage of the on-power each device is consuming. Next, we combine the switching details of a typical device in the 3D-IC with the heat generated obtained as a function of temperature for a single device. We merge our results together in KCL-type thermal network equation (3). Using different numerical methods, including bi-conjugate gradient method, we obtain the thermal map of a 3D-IC in conjunction with individual device operations.

III. SIMULATION RESULT

We tested our methodology on hypothetical digital 3D-ICs, which have each layer modeled after a Pentium III chip with forty million devices [3-4]. Specifically, we obtained temperature maps of a planar chip, and the middle layers of three and six layered 3D-ICs. In Fig. 1, we show our three layered 3D-IC. Our six layered 3D-IC is also stacked in a similar fashion.

In Fig. 3, we show our calculated temperature profiles for the aforementioned 3D-ICs. The temperature reaches more than a hundred degrees above the ambient, which may be detrimental for device and IC operation. The peak temperature value increases from 323 degrees Kelvin in a planar chip to 366 degrees Kelvin in a three layered 3D-IC. It reaches the peak value of 448 degrees Kelvin in a six layered 3D-IC. Likewise the maximum temperature in the bottom layer rises from 323 to 350 to 370 degrees Kelvin as the number of layers increase from one to three to six, respectively. The minimum temperature value also increases from one to six layered 3D-ICs in

the order of 310, 331 and 365 degrees Kelvin. We attribute this to the increase in the net thermal equivalent resistance from each node to ambient. This rise is especially pronounced in 3D-ICs because the thermal resistance of the oxide layers between each stacked chip is very high compared to the lateral thermal resistances, since SiO_2 also acts like a thermal insulator. Thus it traps heat in the middle layers.

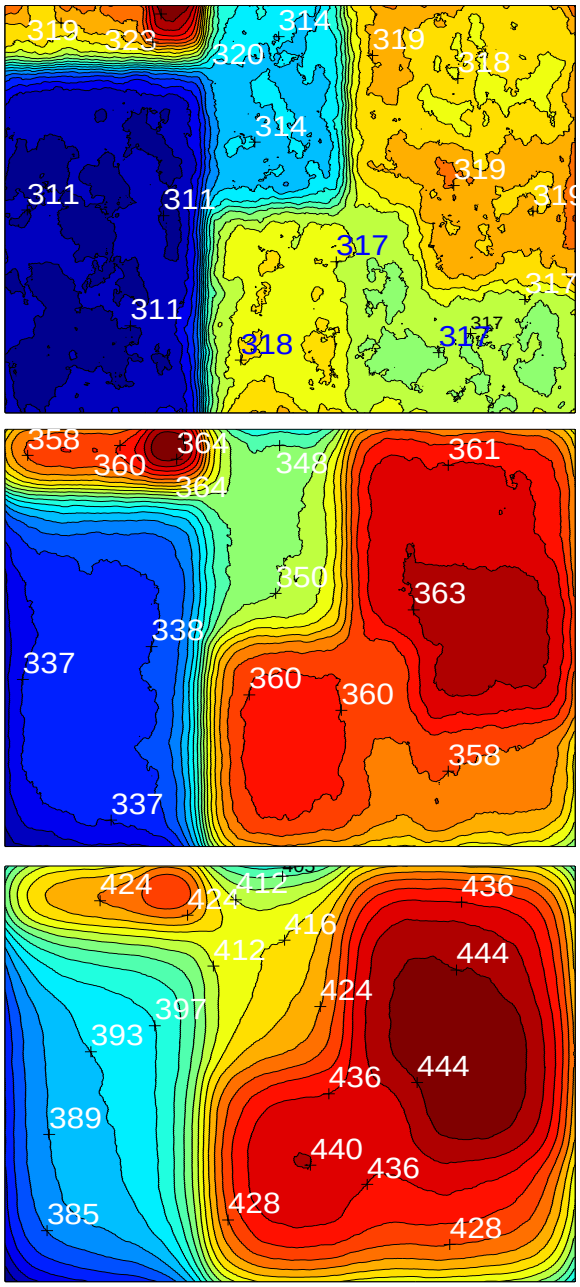

Figure 3: Our calculated temperature maps for a planar chip (top map) and middle layers of three (middle map) and six (bottom map) layered 3D-ICs

Additionally, as the number of layers increase temperature rises proportionately. This shows that thermal resistances of the insulating layers are the main factor that determines the rise in temperature.

In addition, we also note that the location of the peak temperature moves from upper left (corresponding to clock block) to mid-right (corresponding to issue block) as the number of layers increases, which necessitates a different strategy for the 3D-IC design. To test the importance of chip layout, we also simulate our 3D-ICs with each layer rotated by ninety degrees relative to the neighboring layers. This helps to spread out temperature variations and results in a more homogeneous temperature profile in each layer. We also find that the oscillation frequency of a 31 stage ring oscillator decreases by approximately 50% if we change the location from the coolest part of the 3D-IC to the hottest point in the 6-layer 3D chip.

IV. CONCLUSION

We applied our algorithm to 3D-ICs that have each layer modeled after a Pentium III. Our calculated results show that heat flow is blocked strongly by the oxide layers between the stacked chips. Thus middle layer heats up more than the other layers. This may render the circuit and device operation inoperable in these layers. It may also result in permanent structural damage. Thus characterization of the heating problem in 3D-ICs is very important. We provide a model for heating of 3D-ICs in accordance with individual device operations, chip layout, floor plan and average workload. We also can offer solutions to remove the accumulated heat from the center by means of thermal contacts and/or design criteria that are different than those of planar chips.

V. REFERENCES

[1] S S. Im, K. Banerjee, "Full Chip Thermal Analysis of Planar (2D) and Vertically Stacked Integrated (3D) High Performance ICs," *IEDM*, pp. 727-30, 2000.

[2] P. Gelsinger, "Microprocessors for the new Millenium: Challenges, Opportunities, and New Frontiers," *ISSCC*, pp. 22-5, 2001.

[3] A. Akturk, N. Goldsman and G. Metze, "Coupled Modeling of Time-Dependent Full-Chip Heating and Quantum Non-Isothermal Device Operation," *SISPAD*, pp. 311-4, 2003.

[4] http://www.intel.com

[5] S. S. Lee and D. J. Allstot, "Electrothermal Simulation of Integrated Circuits," *Solid-State Circuits*, vol. 28, iss. 12, pp. 1283-93, 1993.