Scheduling Tests for 3D SoCs with Temperature Constraints

Indira Rawat
Department of Electrical Engineering, Govt. Engineering College, Ajmer, Rajasthan , India. .
indirarawat@gmail.com

M.K. Gupta
Deptt. Of Electronics and Communication Engg., Maulana Azad National Institute of Technology, Bhopal, India.
gupta4@yahoo.com

Virendra Singh
Deptt. Of Electrical Engineering, Indian Institute of Technology, Mumbai, India
virendra@computer.org

Abstract

In this paper the test scheduling of 3D SoC has been considered taking the thermal aspect into account. The 3D SoC consists of a complete system stacked vertically. Each stack or layer can have many functional blocks and any floorplan. Test scheduling of the stacked layers has been considered. We have built up a 3D stack for an assumed floorplan. The floorplan for the different layers can be similar or different. The testing of each functional block or core is to be done. Testing will result in production of heat. Heat will spread to its neighboring blocks resulting in their temperature rise. If the testing is done sequentially then temperature rise will be more. Test scheduling is to be done such that the temperature rise does not exceed the limits.

1. Introduction

There has been an enormous advancement in technology over the past decades resulting in improved performance and productivity. 3D integration is considered as a key player in technology improvement in the years to come. The transformation to 3D offers many advantages over 2D including reduced interconnects lengths, better performance and heterogeneous system integration. Many difficulties are also faced in the adoption of 3D, the major being testing issues, thermal constraints and the EDA tools. Test scheduling is one of the major area to be considered in 3D. Much literature is available in 2D test scheduling which cannot be directly applied to 3D [1].

The most important challenge for 3D includes the thermal management. A 3D circuit will have multiple layers of devices and high density of interconnects, consequently several heat generating surfaces or sources Excessive temperature gradients can occur which can cause permanent damage and cause major setback for implementation of this technology.

Origins of 3D date back to 1960 when James Early of Bell Laboratories discussed 3D stacking of electronic components and predicted that heat removal would be the primary challenge to its implementation [6]. Muhannad S. Bakir et al. [7] discussed many unknowns of 3D viz. How to cool? How to deliver power? How to package? Types of intrastratal interconnects, how to assemble? Bond and many more issues. Thermal analysis of vertically integrated circuits was presented for the first time by Michael Kleiner et al [4] in 1995.

Test issues also assume importance in these circuits. Testing generates heat which is to be dissipated at a rate equal or faster than the heat generation. If the heat is not removed permanent damage to the chip may be caused. The other problems of testing are encountered here also mainly the test access, testing time and combined with the thermal problem the issue gets aggravated. TSVs are an important component of these circuits providing power, clock signals and test access. Issues related to TSVs are also to be addressed. They have to be tested for defects and problems arising due to thermal effects. Test time minimization problems are discussed by Z. He, Z. Peng et al in [5] for 2D circuits.

2. Modeling of 3D Structure

It is essential to model the vertical stacks before analysis is carried out. Many models have been proposed [2,4] . An equivalent RC circuit can be used to model the 3D structure. For an integrated circuit at the die level, conduction is the dominant mode which determines temperature. For simplification purpose we can prepare a model of resistors for the circuit we are considering. Each block on the die will have a resistor to its neighbour [3]. This will include neighbours on left, right up and down. For a three dimensional case there will be six neighbours or cells which will be considered.

978-1-4799-2096-9/13/$31.00 ©2013 IEEE

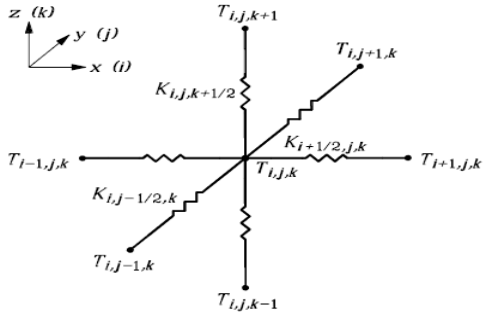

Figure 1. Conductances associated with a single cell

In this paper thermal aware test scheduling of 3D SoCs has been considered. The power dissipated by blocks under test needs to be modeled. The power profile captures the power dissipation of a block over time when applying a sequence of test vectors to inputs and/or pseudo inputs of the blocks. The temperature rise of the chip will be considered while scheduling so that the rise is not above limits leading to hotspot formation.

3. Problem Formulation

Given a stack of n layers consisting of 1, 2… i layers and each layer having 1, 2…n cores and given floorplan. The maximum power constraint P_{max} is given. Find an efficient test sequence such that temperature rise should not exceed the permissible limits and there should be no hotspot formation.

3.1. Heat Flows

There are three modes of heat transfer- conduction, convection and radiation. Conduction is the main mode of heat transfer in solids and this is the mode considered in this work. Heat flow by conduction is governed by Fourier law. $q = -k\frac{dT}{dx}$

This is the one-dimensional heat equation q is the heat flux in W/m^2 , k is the thermal conductivity of the material in (W/m.K). This equation says that heat flux q, (the flow of heat per unit area per unit time) at a point in a medium is directly proportional to temperature gradient at that point. The minus sign indicates that heat flows in the direction of decreasing temperature. If q is written as Q/A where Q is the heat transfer rate, A is the conducting area, and L is the length of material, now equation becomes:

$$Q = -kA\frac{(T_1 - T_2)}{L}$$

i.e temperature drop divided by heat transfer rate, we can write an equivalent expression as

$$\frac{(T_1 - T_2)}{Q} = \frac{L}{kA} = R_{th}$$

Based on this duality between electrical and thermal quantities has been derived [3] viz. Q, the heat transfer rate or power in W is dual of current in A, temperature difference (K) corresponds to voltage difference (V), thermal resistance R_{th} in K/W is dual of R, electrical resistance in ohms, C_{th} the thermal capacitance in J/K is analogous to C, electrical capacitance in F. On similar concepts laws of electrical circuits can be applied to thermal circuits as well. One of the well known laws of electrical is the principle of superposition which states that total electrical current through any branch in a circuit is the algebraic sum of all the currents through that branch. Since duality between electric and thermal circuits has been established , this principle can be applied to thermal circuits as well taking heat flow and temperature rise.

We consider test scheduling of a 3D stack similar to the one shown in Fig.2 where stacks can be built using different floorplans. The cores on the floorplan will have neighbors on all sides. Each core can have a maximum of six neighbors. We have neglected the effect of thermal interface material between the two stacks. When a core is tested, heat is generated which spreads to all its neighbors both vertically and horizontally. The test scheduling approach used in this paper is based on the calculation of total heat produced during a particular testing time. Each vertical stack consists of an SoC which in turn consists of many neighboring blocks.

For three dimensional case consider a cell (i,j,k) with side length Δx_i, Δy_j ,Δz_k . At most there will be 6 adjacent neighbors for each cell in 3D case. The heat flow $Q_{i,j,k+1/2}$ (W) from one cell (i,j,k) to cell above (i,j,k+1) is given by conductance between the two cells multiplied by the temperature difference between the two cells. Similarly heat flow from cell (i,j,k) to other six cells in its proximity are given by following equations.

$$Q_{i,j,k+\frac{1}{2}} = K_{i,j,k+\frac{1}{2}}(T_{i,j,k} - T_{i,j,k+1})(W)$$

$$Q_{i,j,k-\frac{1}{2}} = K_{i,j,k-\frac{1}{2}}(T_{i,j,k} - T_{i,j,k-1})(W)$$

$$Q_{i-\frac{1}{2},j,k} = K_{i-\frac{1}{2},j,k}(T_{i,j,k} - T_{i-1,j,k})(W)$$

$$Q_{i+\frac{1}{2},j,k} = K_{i+\frac{1}{2},j,k}(T_{i,j,k} - T_{i+1,j,k})(W)$$

$$Q_{i,j+\frac{1}{2},k} = K_{i,j+\frac{1}{2},k}(T_{i,j,k} - T_{i,j+1,k})(W)$$

$$Q_{i,j-\frac{1}{2},k} = K_{i,j-\frac{1}{2},k}(T_{i,j,k} - T_{i,j-1,k})(W)$$

$K_{i,j,k+1/2}$ is the conductance between the two cells (i,j,k) and (i,j,k+1) in W/K. The value of conductance is calculated as

$$K_{i,j,k+\frac{1}{2}} = \frac{\Delta x_i \Delta y_j}{\frac{\Delta z_k}{2\lambda_{i,j,k}} + \frac{\Delta z_{k+1}}{2\lambda_{i,j,k+1}}}$$

$\frac{\Delta z_k}{2\lambda_{i,j,k}}$ is the thermal resistance in z direction for half of the cell (i,j,k) $\frac{\Delta z_{k+1}}{2\lambda_{i,j,k+1}}$ is the thermal resistance for half of the cell (i,j,k+1). $\lambda_{i,j,k}$ is the thermal conductivity in W/mK .

Similarly there are five other thermal conductances associated with each cell [2] which can be similarly computed. The conductances between two cells (i,j,k) and (i,j,k+1) is shown in Figure. 1.The heat flow is calculated based on electrical equivalents of the thermal circuits and are based on [2]. The changes for the boundary cells can also be suitably done. Though the heat spread cannot be restricted to 6 neighbors alone, they will again spread to their adjacent neighbors and some cooling will also be there but this has been assumed negligible here.

The total heat flow to cell (i,j,k) from six neighboring cells can be calculated. As a core is tested it generates heat. The heat flow to its neighboring core can be calculated using the equations mentioned above. The heat flow results in rise or fall of temperature. The new temperature is given by

$T^{new} = T^{old} + \Delta T$

Principle of superposition has been applied for the purpose of calculating the temperatures of cores.

4. Test Scheduling

The testing of cores results in rise of temperature. Once a core is tested, its impact on all neighboring cores is observed. Next a cool core is selected for testing. Again heat transfer and temperature rise is noted for each core. The new temperature will be the sum of old and the incremental temperatures. The least heated core is selected for testing. Values of temperatures are updated. Another temperature rise is observed with core testing being done sequentially i.e. according to the numbering viz. core $1, 2, 3 \ldots$ n of one layer, then 1,2 3… m of the second layer and so on and temperature rise is observed with each core.

4.1. Algorithm

```
Input= initial heat generated
Output= Sequence of cores for testing, power dissipated
final temperature.
Initial temperature = 300 kelvin
Sequence= 0
Last core = N
l= total number of cores
For i= 1:N
      Read details of each layer
        ( x-coordinate, y coordinate width and height of
        each core)
calculate conductance
end for
 Schedule list = null
Select one of cores for testing and add to schedule list
 While l≠ 0
   Calculate
      Heat generated, temperature rise in neighboring cores,
      Compare with Pmax, update values
Select the core with minimum temperature rise
Test and delete from list
Test calculate heat spread in all cores
Repeat
Till l=0   (All cores are done)
```

4.2. Implementation

The algorithm was implemented for 2, 3, 4 layer structure with the floorplans as shown in Figure. 2. Different combinations of the two floorplans were used to build up the stack. The input to the algorithm is the initial value of heat generated and the output is the sequence of cores for testing, power dissipated and the final temperature. Initial temperature is taken as 300 Kelvin. Details of each layer are read viz. x-coordinate, y-coordinate, width and height of each core in it. Conductance value is calculated for each core both in horizontal and vertical direction to account for the heat spread. One of the cores is selected for scheduling. Heat generation is calculated with some initial value (= 1W) and temperature rise is observed. The heat will spread to its neighbours both horizontally and vertically. Temperature is updated for all. Next the core selected is the one with least temperature rise. It is scheduled next Again heat spread due to this is calculated. The temperature values are updated by adding up the initial temperature and incremental temperatures. It is repeated till all cores are scheduled. P_{max} has been taken as 4 units in this example.

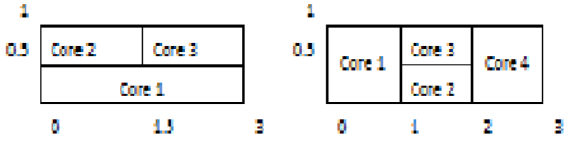

Figure 2. Floorplan of different stacks

5. Results and Conclusion

Temperature rise observed when algorithm is implemented is shown in Table1.

Table 1. Temp. after scheduled and seq. testing

2 layer stack		3 layer stack		4 layer stack	
Temperture(K)		Temperature(K)		Temperature(K)	
Seq.	Schd.	Seq.	Schd.	Seq.	Schd.
346	346.5	346	346.5	347	345
348	346	346	345	347	345
348.5	346	347	345	365	350
355	355	365	350	355	355
360	350	360	355	340	340
365	357	355	355	385	365
355	355	340	340	370	375

The graph in Figure. 3, Figure. 4 and Figure. 5 shows the comparison for 2,3 and 4 stack structures respectively for sequential and scheduled testing. The two sets of temperatures are plotted on the same graph to compare the final temperature of the two methods of testing. The peak temperatures obtained for the two cases show a fair difference. The goal of the algorithm was to avoid peak temperatures in order to avoid hotspot formation, which has been achieved- the peak temperatures differ by 10-15 K which shows a fair control over temperature.

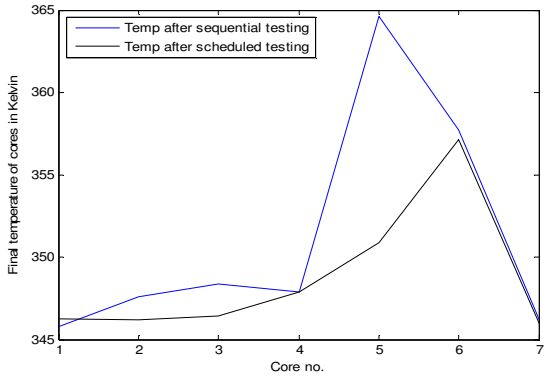

Figure.3. Comparison of temp. of cores for 2 stacks

The graphs also indicate that the temperature of innermost core is the highest viz. for cores 4,5,6,7 etc. as there is no way for the heat to escape, a point of concern in 3D circuits. This also calls for some cooling schemes to be considered which is the focus for our future work. The algorithm can be further implemented for some benchmark circuits. For further work in this, time minimization will be considered along with thermal aware test scheduling and implementation on some benchmark circuits. Parallel testing of cores with same temperature rise can be considered.

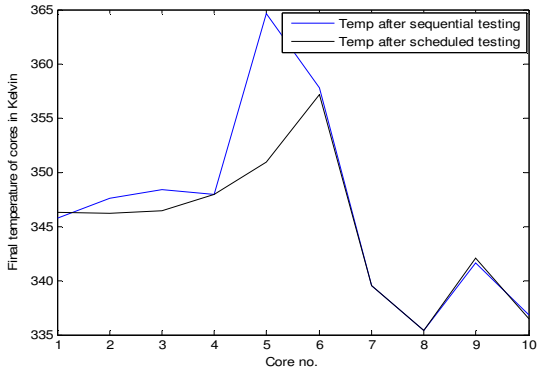

Figure. 4 Comparison of temp. of cores for 3 stacks

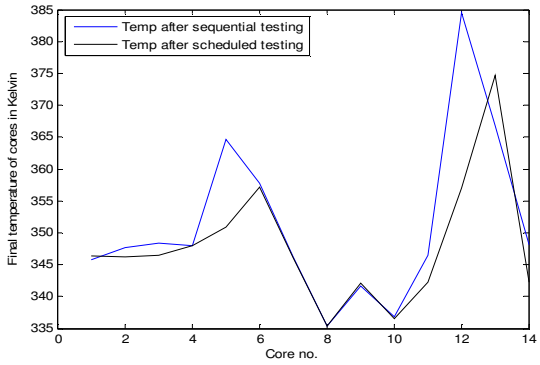

Figure 5 Comparison of temp. of cores for 4 stacks

References

[1]. Paul Rosinger, Bashir M. Al-Hashimi, "Thermal-Safe Test Scheduling for Core-Based System-on-Chip Integrated Circuits", *IEEE Transactions on Computer Aided Design*, vol 25, no. 11, November 2006, pp 56-61.

[2]. Thomas Blomberg, "Heat conduction in two and three dimensions", *Report TVBH* ,1996.

[3]. W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan, "Hotspot: A Compact Thermal Modeling Methodology for Early-Stage VLSI Design", *IEEE Transactions on VeryLarge Scale Integration (VLSI) Systems,* 14(5): May 2006, pp 501–513.

[4] Michael B. Steiner, Stefen A. Kuhn, Peter Ramm and Werner Weber, "Thermal Analysis of Vertically Integrated Circuits", *Proc. IEDM 1995*, pp 487-490.

[5] Z. He, Zebo Peng and Eles P, "A Heuristic for Thermal-safe SoC Test Scheduling", *IEEE International Test Conf,.*Oct. 2007, pp 1-10.

[6] J. Early, "Speed, Power and Component Density in Multielement High Speed Logic System", *Proc. IEEE International Solid State Circuits Conf,* Feb. 1960, pp 78-79.

[7] Muhannad S. Bakir, Calvn King et al, " 3D Heterogeneous Integrated Systems: Liquid Cooling, Power Delivery and Implementation" *IEEE Custom Integrated Circuits Conf, (CICC)* ,Oct. 2008, pp 663-670.