Physical Design Issues in 3-D Integrated Technologies

Vasilis F. Pavlidis, Eby G. Friedman

Physical Design Issues in 3-D Integrated Technologies

Vasilis F. Pavlidis[1] and Eby G. Friedman[2]

[1]LSI EPFL, 1015 Lausanne, Switzerland
[2] Department of Electrical and Computer Engineering, University of Rochester,
Rochester, New York 14627, USA
vasileios.pavlidis@epfl.ch, friedman@ece.rochester.edu

Abstract. Design techniques for three-dimensional (3-D) ICs considerably lag the significant strides achieved in 3-D manufacturing technologies. Advanced design methodologies for 2-D circuits are not sufficient to manage the added complexity caused by the third dimension. Consequently, design methodologies that efficiently handle the added complexity and inherent heterogeneity of 3-D circuits are necessary. These 3-D design methodologies should support robust and reliable 3-D circuits, while considering different forms of vertical integration, such as systems-in-package and 3-D ICs with fine grain vertical interconnections. The techniques described in this chapter address important physical design issues and fundamental interconnect structures in the 3-D design process.

1. Introduction

Technology scaling and CMOS technologies have steadily supported an increase in the performance of integrated circuits (ICs) over the past several decades. These driving forces are expected, however, to lose momentum as the fabrication of nanoscale devices at gigascale densities become increasingly difficult and economically infeasible [1]. Three-dimensional (3-D) integration is a novel design paradigm with great potential to fundamentally advance the computational power and functionality of modern integrated systems [2].

The inherent advantage of 3-D integration is the drastic decrease in interconnect length, particularly the long global interconnects, which directly results in increased speed [3], [4], [5]. The interconnect power is also reduced as the capacitance of the wires decreases [6]. Another characteristic of 3-D ICs of even greater importance than the decrease in the interconnect length is the ability of these systems to include disparate technologies, greatly extending the capabilities of modern systems-on-chip (SoC) [7].

This defining feature of 3-D ICs offers unique opportunities for highly heterogeneous and sophisticated systems [8]. This heterogeneity, however, greatly complicates the interconnect design process within a multi-plane system, as potential design methodologies need to manage the diverse interconnect impedance characteristics and process variations caused by the different fabrication processes and technologies employed in the multiple physical planes [9]. Additional primary

challenges in 3-D circuits include the development of methodologies at the front end of the design process [10], [11], multi-plane functional testing [12], thermal management techniques [13], and maturing manufacturing technologies [14].

Physical and interconnect design techniques for 3-D circuits are the main focus of this chapter. A short description of vertical interconnects in 3-D circuits is offered in the following section, demonstrating the diverse characteristics of this revolutionary design paradigm. These traits, in turn, pose new constraints and requirements on the physical and interconnect design process. The primary physical design issues, namely floorplanning, placement, and routing for 3-D ICs, are discussed in Sections 3, 4, and 5, respectively. An approach to place the vertical interconnects to decrease the interconnect delay is described in Section 6. The important task of synchronization is considered in Section 7. Experimental results from a 3-D test circuit are also presented. The concept of 3-D NoC for improving the communication throughput within a system-on-chip while reducing interconnect design complexity is presented in Section 8. Several topologies and related improvements in the speed and power consumed by these global interconnects are also described in this section. The key points of this chapter are summarized in Section 9.

2. Vertical Interconnects

There are multiple ways to vertically interconnect 3-D circuits. The characteristics of the different vertical interconnects and the requirements associated with this type of interconnect structure are discussed in this section. To exemplify the role of these interconnects, consider the 3-D circuit shown in Fig. 1. Two different types of interconnect can be distinguished in Fig. 1. The horizontal or intraplane interconnects connecting circuits located within the same plane and the interplane interconnects connecting circuits located on different planes. The interplane wires comprise horizontal and vertical segments.

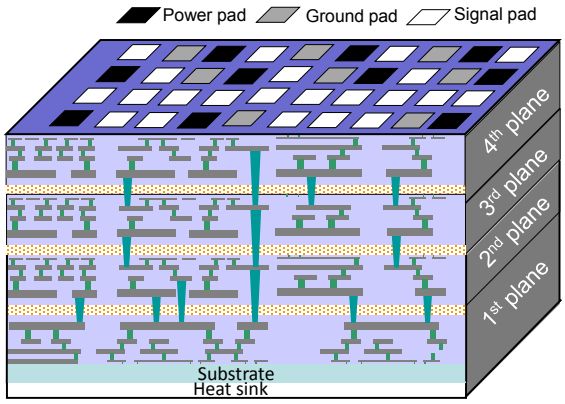


Fig. 1. Schematic of a 3-D IC consisting of four planes.

The interconnects through the *z-axis* (*i.e.,* vertical) can be implemented with several means, such as solder balls, wire bonds, and vertical interconnects that are etched through the silicon substrate. The latter type of interconnects is typically called

a through silicon via (TSV) [15]-[18]. The density of the vertical interconnect dictates the granularity of the interconnected planes of the 3-D system, directly affecting the interplane communication bandwidth.

Coarsely interconnected 3-D systems include several either bare or packaged dice connected along the third dimension which are typically described as a system-in-package (SiP) [17]. The predominant benefits of SiP are the increased packaging efficiency as compared to 2-D integrated systems and shorter off-chip interconnects. The deleterious effects of the long on-chip interconnects, however, are not mitigated. These issues are effectively resolved by another form of vertical integration, called simply (and somewhat abstractly) 3-D ICs.

Three-dimensional circuits can be conceptualized as the bonding of multiple wafers or bare dice. The distinctive difference between an SiP and a 3-D IC is the granularity of the vertical interconnects. Examples of 3-D systems connected with different means are illustrated in Fig. 2. In addition to the different types of vertical interconnects, several bonding styles for 3-D ICs are also possible: front-to-front, back-to-front, and back-to-back are some of these styles which are also depicted in Fig. 2.

Other important criteria related to manufacturing TSVs include the reliability and cost of these structures. A high TSV aspect ratio, the ratio of the diameter of the top edge to the length of the via, may also be required for certain types of 3-D circuits. The effect of forming the TSVs on the performance and reliability of neighboring active devices should also be negligible.

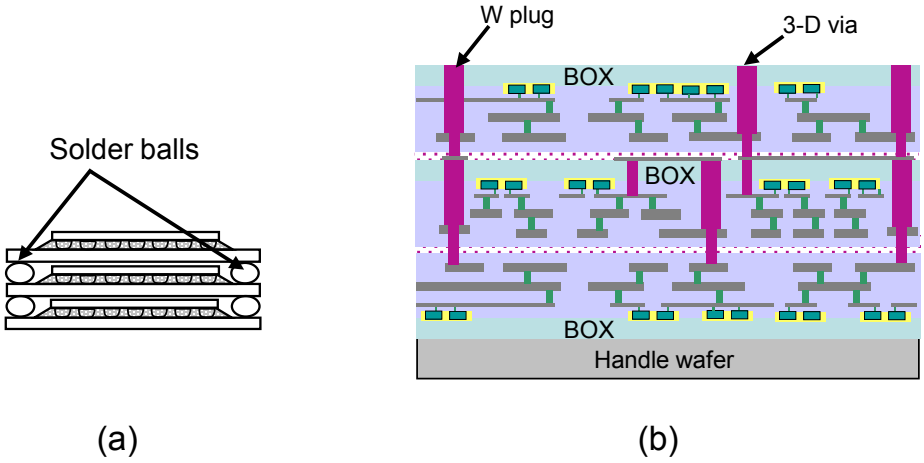


Fig. 2. Different forms of 3-D integration (not to scale), (a) system-in-package (SiP) [17] and (b) a 3-D circuit with dense through silicon vias [20], [21]. Two different bonding styles, front-to-front and front-to-back, are illustrated. The W plug is composed of tungsten.

Furthermore, producing TSVs with low impedance characteristics is another primary goal of 3-D manufacturing technologies since these characteristics can degrade the performance benefits that stem from the decreased wirelength in 3-D circuits. Finally, not properly characterizing the contribution of the TSVs to the overall delay of the critical interplane interconnect can result in significant inaccuracy

in estimating the performance of a 3-D system [19]. Consequently, these structures must be carefully considered during the 3-D physical design process. Examples of a TSV used in CMOS and SOI circuits are illustrated in Figs. 3a and 3b, respectively. The impedance and physical characteristics of these structures are listed in Table 1. A pitch equal to twice the diameter of the TSV is assumed where this dimension is not provided.

The thermal traits of the TSVs are also significant as these vias can affect the thermal behavior of a 3-D IC. TSVs can be used to provide high thermal conductivity paths to facilitate the flow of heat from the upper planes to the plane attached to the heat sink, maintaining the temperature of a 3-D circuit within acceptable levels. Since the vertical interconnects affect the performance of 3-D systems, the treatment of these interconnects is central to the development of 3-D physical and interconnect design techniques. The most important steps of the 3-D design process and related design methodologies are discussed in the remainder of this chapter.

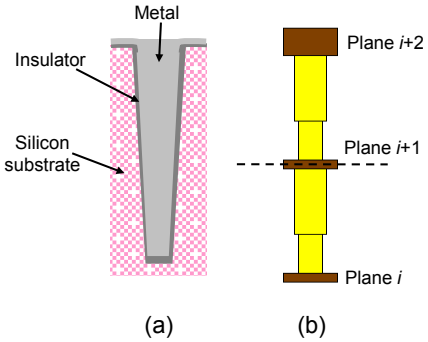


Fig. 3. Examples of a through silicon via (not to scale) used in (a) SiP and 3-D CMOS technologies [14], [22] and (b) 3-D SOI processes [21].

Table 1. Impedance and physical characteristics of TSVs.

Process	*Depth [μm]*	*Diameter [μm]*	*Total resistance [mΩ]*	*Density [1/mm^2]*
[22]	25	4	140	~1.6×10^4
[11]	30	1.2	<350	~1.7×10^5
[23]	90	75	2.4	~44
[21], [24]	~12	1.75	148	~8.2×10^4

3. Floorplanning for 3-D Circuits

The predominant design objective for floorplanning a circuit has traditionally been to achieve the minimum area or, alternatively, the maximum packing density while interconnecting these blocks with minimum length wires. Most floorplanning algorithms can be classified as either slicing [25] or non-slicing [26]. Floorplanning techniques belonging to both of these categories have been proposed for 3-D circuits [27]-[29]. An efficient floorplanning technique for 3-D circuits should adequately handle two important issues; representation of the third dimension and the related

increase in the solution space. Floorplanning techniques for 3-D circuits that address these issues are discussed in this section. Multi-objective techniques are also reviewed.

Notating the location (*i.e.,* the *x, y, z,* coordinates) and dimensions (*i.e.,* width, length, and height) of the circuit cells in a volumetric system typically requires a considerable amount of storage. A 3-D circuit, however, consists of a limited number of planes. Consequently, such a system can be described as an array of two-dimensional planes, where circuit cells are treated as rectangles that can be placed on any of the planes within a 3-D system [13], [28], [29]. The second challenge for 3-D floorplanning is to effectively explore the solution space, where a hierarchical approach can often be more efficient for floorplanning 3-D circuits than a flat approach.

In non-hierarchical floorplanning algorithms, the floorplanning process proceeds by assigning the cells to the planes of the stack followed by simultaneous intraplane and interplane cell swapping, potentially exploring the entire solution space. Interplane moves, however, result in a formidable increase in the solution space, directly affecting the computational time of a flat floorplanning algorithm.

Alternatively, a hierarchical approach can be used to significantly reduce the number of candidate solutions, where a two step solution to the floorplanning problem is followed. Initially, the circuit cells are assigned to the physical planes. In the second step, a simulated annealing based engine simultaneously generates the floorplan of each of the planes by only permitting intraplane moves, considerably decreasing the search space for the optimal floorplan [28]. An example of the increase in the solution space due to the third dimension is illustrated in Figs. 4a and 4b.

The partitioning scheme adopted in the initial step of the hierarchical approach plays a crucial role in determining the compactness of a particular floorplan, as interplane moves are not allowed when floorplanning the planes. Different partitions correspond to different subsets of the solution space which may exclude the optimal solution(s). The criterion for partitioning should therefore be carefully selected. Partitioning can, for example, be based on minimizing the estimated total wirelength of the system [30] and/or the number of vertical interconnects [31]. Application of a hierarchical approach to the MCNC and GSRC benchmark suites [32] demonstrates a small reduction, on the order of 3%, in the number of vertical vias and a significant 14% reduction in wirelength, as compared to non-hierarchical 3-D floorplanning techniques [13], [30], [31].

The complexity of three-dimensional integration requires several dissimilar metrics for producing efficient floorplans for 3-D circuits beyond the use of traditional area and wirelength metrics. These metrics can consider, for example, communication throughput among the circuit blocks [33] or the number of interplane vias [13]. Techniques that include a thermal objective have also been developed [13]. The thermal objective typically aims at producing a uniform temperature distribution across each plane while peak temperatures are maintained sufficiently low. A multi-objective cost function inevitably increases the total computational runtime. A significant portion of this time is attributed to thermal profiling the 3-D circuit each time a candidate floorplan is generated. To reduce this time, simple thermal models are utilized, slightly degrading the quality of the solution [13].

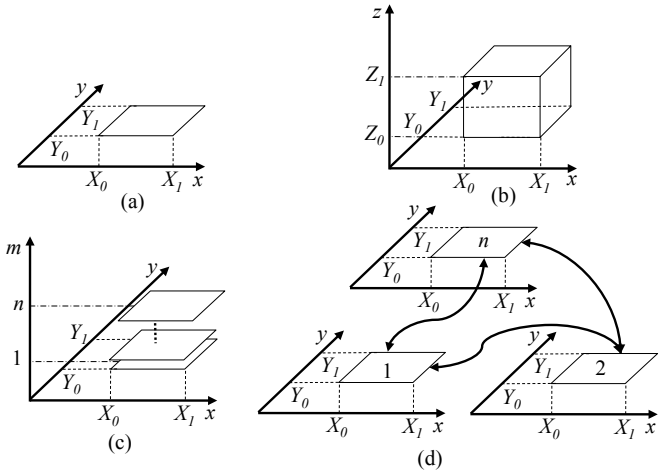


Fig. 4. Example of physical design solution space for floorplanning 2-D and 3-D circuits, (a) available area for floorplanning a planar circuit, (b) available volume for floorplanning a 3-D circuit, (c) a finite number of planes is considered to reduce the solution space, and (d) the floorplan of the planes is generated after the circuit cells are assigned to each plane. The arrows represent global constraints among planes that guide the floorplan of a 3-D system.

4. Placement for 3-D Circuits

Placement algorithms have traditionally targeted minimizing the area of a circuit and the interconnect length among the cells, while reserving space for routing the interconnect. In vertical 3-D integration, a "placement dilemma" arises in deciding whether two circuit cells sharing a large number of interconnects can be more closely placed within the same plane or placed on adjacent physical planes, decreasing the interconnection length. Placing the circuit blocks on adjacent planes can often produce a line with the shortest wirelength to connect these blocks. An exception is the case of small blocks within an SiP where the length of the interplane vias is greater than 100 μm [34]. Placement methodologies have also been discussed where other objectives, such as thermal gradients among the physical planes and the temperature of the planes [35], are considered.

Several approaches have been adopted for placing circuit cells within a volume [29], [36]. Different types of circuit cells for various 3-D technologies have been investigated in [37]. Layout algorithms for these cells have also been devised, demonstrating the benefits of 3-D integration. Since TSVs consume silicon area, possibly increasing the length of some interconnects, an upper bound on this type of interconnect resource is necessary. Alternatively, sparse utilization of the vertical interconnects can result in an insignificant savings in wirelength. To consider the effect of the vertical interconnects, a weighting factor can be used to increase the distance in the vertical direction, controlling the decision as to where to insert the

interplane vias [38]. This weight essentially behaves as a controlling parameter that favors the placement of highly interconnected cells within the same or adjacent physical planes.

Alternatively, TSVs can be treated as circuit cells since these interconnects occupy silicon area [39] and are included in the individual cell placement process within each plane as illustrated in Fig. 5. Although these approaches consider the location of the TSV, the fundamental objective is to decrease the interconnect length. The maximum achievable reduction in the interconnect length for the longest on-chip interconnect is proportional to $\sqrt{n}$ where n is the number of planes constituting a 3-D system [6]. Any further improvement in the performance of the interplane interconnects can be obtained by considering the electrical characteristics of the TSV. A placement methodology that exploits these characteristics is discussed in Section 6.

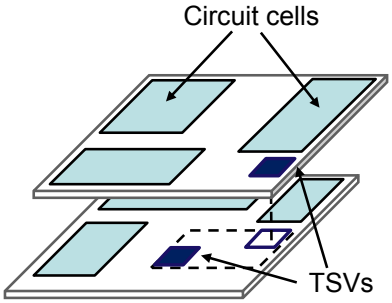


Fig. 5. Treating the TSVs as circuit cells on different planes can result in two different locations for placing a TSV. These locations define a region in which the TSV can be placed to satisfy different design objectives.

As with floorplanning, multi-objective placement techniques for 3-D circuits are necessary. Additional objectives that affect both the cell placement and wirelength are simultaneously considered. The force directed method is a well known technique used for cell placement [40], where repulsive or attractive forces are placed on the cells as if these cells are connected through a system of springs. The force directed method has been extended to incorporate the thermal objective during the placement process [41]. In this approach, repulsive forces are applied to those blocks that exhibit high temperatures (*i.e.,* "hot blocks") to ensure that the high temperature blocks are placed at a greater distance from each other. The efficiency of this force directed placement technique has been evaluated on the MCNC [42] and IBM-PLACE benchmarks [43], demonstrating a 1.3% decrease in the average temperature, a 12% reduction in the maximum temperature, and a 17% reduction in the average thermal gradient. The total wirelength, however, increases by 5.5%.

Alternatively, additional TSVs that do not function as a signal path can be utilized to further enhance the heat transfer process. The design objective is to identify those regions where thermal vias are most needed (the hot spots) and place thermal vias within those regions at the appropriate density. Such an assignment, however, is mainly restricted by two factors; the routing blockage caused by these vias and the size of the unoccupied regions or white space that exist within each plane. Although

thermal via insertion can be applied as a post placement step, integrated techniques produce a more efficient distribution of the thermal TSVs for the same temperature constraint [30]. The integrated technique requires 16% fewer thermal vias for the same temperature constraint, with a 21% increase in computational time and an almost 3% reduction in total area.

5. Routing for 3-D Circuits

Routing is the most complex and least developed of the physical design techniques used in 3-D circuits. The multiple metal layers available for routing on each physical plane exacerbate the difficulty in routing a net connecting several cells located on different planes. As these interconnects also compete with the transistors for silicon area, routing is a formidable task for 3-D circuits. Early results on routing 3-D circuits demonstrated several issues related to this physical design task [44]. Consequently, several heuristics have been developed that address routing in the third dimension [45], [46].

An effective approach for routing 3-D circuits is to convert the routing interplane interconnect problem into a 2-D channel routing task, as the 2-D channel routing problem has been efficiently solved [47]. A number of methods can be applied to transform the problem of routing the interplane interconnects into a 2-D routing task, which requires utilizing a portion of the available routing resources for interplane routing (usually the top metal layers).

Alternatively, multi-level algorithmic techniques [48] have been applied to route 3-D circuits. The advantages of multi-level routing are the lower computational time and higher completion rates as compared to flat and hierarchical routers. Multi-level routing can be treated as a three stage process, as illustrated in Fig. 6; a coarsening phase, an initial solution generation at the coarsest level (level p) of the grid, and a subsequent refinement process until the finest level of the grid is reached. Before the coarsening phase is initiated, the routing resources in each unit block of the grid are determined by a weighted area sum model. The routing resources are allocated during each coarsening step. The resources for the local nets within a block are transferred at each coarsening step. At the coarsest level, an initial routing tree is generated. This initial routing task commences with a minimum spanning tree for each multi-terminal net. A Steiner tree heuristic and a maze searching algorithm generate a 3-D Steiner tree for each of these interconnects. Additionally, the TSVs are estimated for each block. During the last phase, the initial routing tree is refined until the finest level is reached. In this refinement phase, the signal (and thermal) TSVs are successively assigned and distributed within each block. The routing of the wires follows the refinement of the TSVs. At the finest level, a detailed router completes the routing of the circuit [48].

Although this technique offers a routing solution for standard cell and gate array circuits, alternative techniques that support different forms of vertical integration, for example, systems-on-package (SOP), are also required. In an SOP, the routing problem can be described as connecting the I/O terminals of the blocks located on the planes of the SOP through interconnect and pin layers. For systems where the routing

resources, such as the number of pin distribution layers, are limited, multi-objective routing is required to achieve a sufficiently small form factor [46].

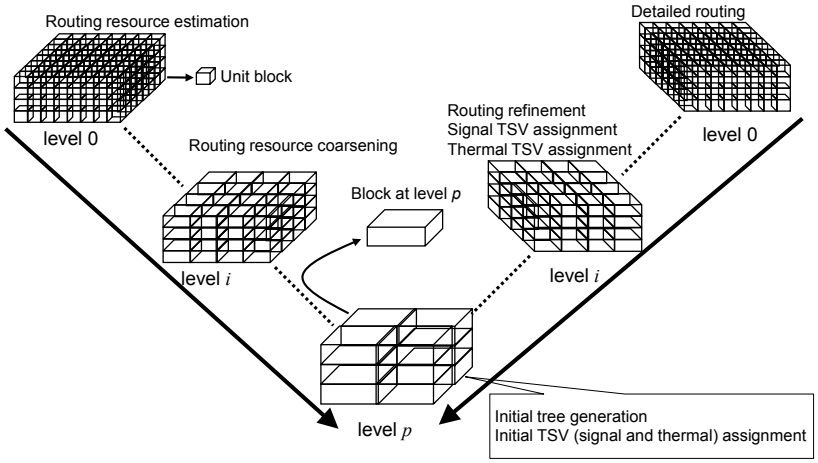


Fig. 6. Multi-level routing for 3-D circuits. The technique can be adapted to include multiple objectives for routing a 3-D circuit [48].

Multi-level routing for 3-D ICs has been extended to include the thermal objective [49]. In addition to routing resources, the power density within each block of the grid is determined at each coarsening step. An initial TSV assignment to each block is implemented during the coarser step along with generation of an initial routing tree. The TSV assignment includes both signal and thermal TSVs, with priority given to the signal TSVs. Alternatively, thermal TSVs are assigned to a block after insertion of the signal TSVs without exceeding the maximum TSV capacity of the block.

6. Timing Optimization of Interplane Interconnects

Three-dimensional integration demonstrates many opportunities for heterogeneous SoCs [9]. Integrating circuits from diverse fabrication processes into a single multi-plane system can result in substantially different interconnect impedance characteristics of each physical plane within a 3-D circuit. By considering the disparate interconnect impedance characteristics of 3-D circuits, the performance of the interplane interconnects can be significantly improved. An efficient technique to decrease the delay of interplane interconnects by optimally placing the TSVs is discussed in this section.

The interplane interconnects are modeled as an assembly of horizontal interconnect segments with different impedance characteristics connected by interplane vias where each segment is modeled as a distributed *RC* line. A schematic of an interplane interconnect connecting two circuits located *n* planes apart is illustrated in Fig. 7. The horizontal segments of the line are connected through the vias, which can traverse more than one plane where each via is placed within a certain physical interval. The via placement is constrained,

$$0 \le x_j \le \Delta x_j , \tag{1}$$

where Δx_j is the length of the interval where the via connecting planes j and j+1 can be placed. This interval length is called the "allowed interval" here for clarity. x_j is the distance of the via location from the edge of the allowed interval.

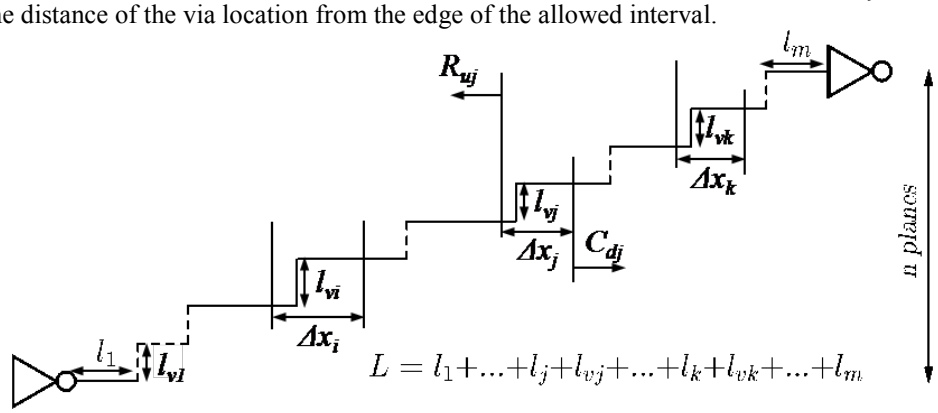


Fig. 7. Interplane interconnect connecting two circuits located n planes apart.

A heuristic for near-optimal interplane via placement of two-terminal nets that include several TSVs has been developed [50]. Based on the Elmore delay model [51], the key concept in the heuristic is that the optimum via placement depends primarily upon the size of the allowed interval (that is estimated or known after an initial placement) rather than the exact location of the via.

This heuristic has been used to implement an algorithm that exhibits an optimal or near-optimal TSV placement for two-terminal interplane interconnects in 3-D ICs and has been applied to relatively short interconnects (< 2 mm) [50]. For these wires, SPICE delay simulations demonstrate an average improvement of 8.9% as compared to the case where the vias are placed at the center of the allowed intervals and 14.1% as compared to random via placement, respectively. The two-terminal via placement algorithm is also compared both in terms of optimality and efficiency to an optimization solver, YALMIP [52]. The algorithm exhibits high accuracy as compared to YALMIP independent of the number of planes that comprise the 3-D interconnect and exhibits a maximum error of 0.01%. Furthermore, the algorithm is approximately two orders of magnitude faster than YALMIP while the complexity of the algorithm exhibits an almost linear dependence on the number of interplane vias.

The two-terminal heuristic can also be adapted to the important class of multi-terminal nets. A simple interplane interconnect tree (also called an interconnect tree for simplicity) is illustrated in Fig. 8. The leaves of the tree are located on different physical planes within a 3-D stack. Sub-trees not directly connected to the interplane vias which do not contain any interplane vias (*i.e.,* intraplane trees) are also shown. The weighted summation of the distributed Elmore delay of the branches of an interconnect tree is considered as the objective function,

$$T_w = \sum_{\forall s_{pq}} w_{s_{pq}} T_{s_{pq}} , \tag{2}$$

where w_{spq} and T_{spq} are the weight and distributed Elmore delay of sink s_{pq}, respectively. Weights are assigned to the sinks according to the relative criticality of the sinks. The constrained optimization problem for placing a via within an interplane interconnect tree can be described as

$$\text{(P1) } \textbf{minimize } T_w, \textbf{ subject to } (1), \forall \text{ via } v_j. \quad (\mathbf{3})$$

The heuristic and related algorithm that solve (3) have been applied to interconnect trees for two different 3-D technologies. These case studies include a 3-D IC technology based on [20] where the TSV length is $l_{v3\text{-}D}$ = 10 μm and an SiP technology where the TSV length is l_{vSiP} = 70 μm [23]. The impedance characteristics of the TSVs are $r_{v3\text{-}D}$ = 22 Ω/mm and $c_{v3\text{-}D}$ = 210 fF/mm and r_{vSiP} = 22 Ω/mm and c_{vSiP} = 6 pF/mm for the 3-D IC and SiP technology, respectively. The savings in delay achieved by optimally placing the vias is listed in Table 2 for different via placement scenarios.

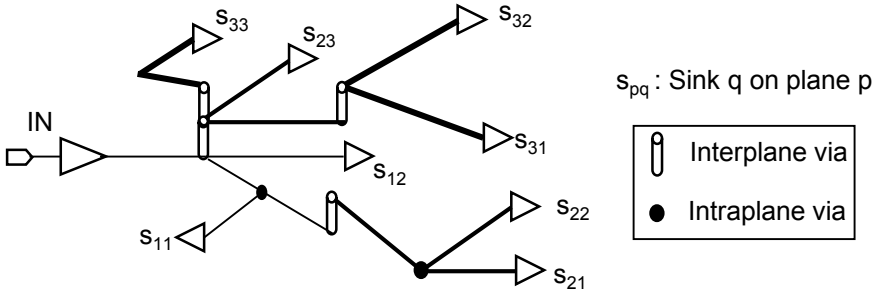


Fig. 8. An example of an interplane interconnect tree.

The improvement in delay of the interconnect trees is listed in columns 6 through 9 of Table 2. The results are compared to the case where the vias are initially placed at the center of the allowed interval (*i.e.,* $x_i = \Delta x_i/2$) and the case where the vias are placed at the lower edge of the allowed interval (*i.e.,* $x_i = 0$). The improvement in delay depends upon the length of the allowed interval. Note that the improvement in delay achieved by optimally placing the TSVs in a 3-D IC is substantially greater than the improvement for an SiP technology. This difference is due to the significantly longer length and larger impedance characteristics of the TSVs utilized in an SiP. Manufacturing processes that provide short vertical interconnects with low parasitic impedances are therefore necessary; otherwise, the performance benefits due to the reduction in interconnect length will decrease since the TSVs contribute significantly to the overall interconnect delay.

From these results, exploiting the non-uniform impedance characteristics of the interplane interconnects when placing the vias can improve the delay of multi-plane lines. This improvement in delay can decrease the number of repeaters required to drive a global line or eliminate the need for repeaters in semi-global (medium length) lines. In addition, wire sizing can be avoided, thereby saving significant power. Decreasing the number of repeaters and avoiding wide lines reduce the overall power consumption, which is a particularly important issue in 3-D circuits.

Table 2. Delay of various interplane interconnect trees for different number of sinks, physical planes *n*, and 3-D technologies.

n	*Technology*	*Number of sinks*	*Avg. branch length [μm]*	*Δx_i's [μm]*	*Delay improvement [%]* $x_i^* = \Delta x_i/2$		$x_i^* = 0$		*Instances*
					Avg	*Max*	*Avg*	*Max*	
3	3-D IC	4	216	50	1.31	7.11	5.33	13.00	10000
4	3-D IC	8	407	50	1.47	6.88	6.83	13.22	10212
3	3-D IC	4	815	150	1.15	5.74	4.42	10.02	11000
4	3-D IC	8	909	150	1.29	4.98	5.70	9.48	10219
3	SiP	4	216	50	1.21	4.99	1.78	5.58	10000
4	SiP	8	407	50	0.90	3.54	1.98	5.72	10212
3	SiP	4	815	150	1.31	4.10	1.98	5.68	11000
4	SiP	8	909	150	1.04	3.28	2.34	5.71	10219

7. Synchronization in 3-D Circuits

An omnipresent and challenging issue for synchronous digital circuits is the reliable distribution of the clock signal to the many thousands of sequential elements distributed throughout a synchronous circuit [53], [54]. The complexity is further increased in 3-D ICs as sequential elements belonging to the same clock domain (*i.e.,* synchronized by the same clock signal) can be located on different planes. Another important issue in the design of clock distribution networks is low power consumption, since the clock network dissipates a significant portion of the total power consumed by a synchronous circuit [55]. This demand is stricter for 3-D ICs due to the increased power density and related thermal limitations.

In 2-D circuits, symmetric interconnect structures, such as H- and X-trees, are widely utilized to distribute the clock signal across a circuit [54]. The symmetry of these structures permits the clock signal to arrive at the leaves of the tree at the same time, resulting in synchronous data processing. Maintaining this symmetry within a 3-D circuit, however, is a difficult task. Consequently, asymmetric structures are useful candidates for distributing the clock signal within a 3-D circuit. Issues related to the distribution of the clock signal within a 3-D system are discussed in this section. Experimental results of a 3-D test circuit manufactured by MIT Lincoln Laboratories composed of several different 3-D clock network architectures are also described.

To evaluate the specific requirements of a 3-D clock network, consider a traditional H-tree topology. At each branch point of an H-tree, two branches emanate with the same length. An extension of an H-tree to three dimensions does not guarantee equidistant interconnect paths from the root to the leaves of the tree. Note that the vertical interconnects are of significantly different length as compared to the horizontal branches and exhibit different impedance characteristics.

A test circuit exploring four different clock network topologies for 3-D circuits has been designed, manufactured, and measured. The test circuit is based on a 3-D fully depleted silicon-on-insulator (FDSOI) fabrication technology recently developed by MIT Lincoln Laboratories (MITLL) [20]. The MITLL process is a wafer level 3-D integration technology with up to three FDSOI wafers bonded to form a 3-D circuit. The minimum feature size of the devices is 180 nm, with one polysilicon layer and

three metal layers interconnecting the devices on each wafer. A backside metal layer also exists on the upper two planes, providing the starting and landing pads for the TSVs, and the I/O, power supply, and ground pads for the entire 3-D circuit.

Each block contains the same logic circuit with different clock distribution networks. The off-chip clock signal is received by the clock driver through an RF pad located at the middle of each block. Additional RF pads are placed at different locations on the topmost plane of each block for probing. The fabricated test circuit is depicted in Fig. 9, where the RF and DC pads on the back side metal layer of the third plane are shown.

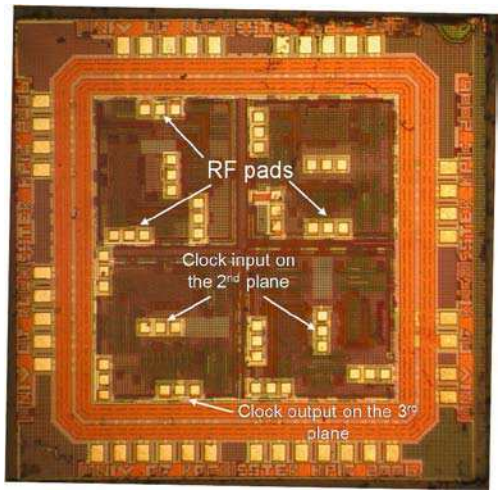
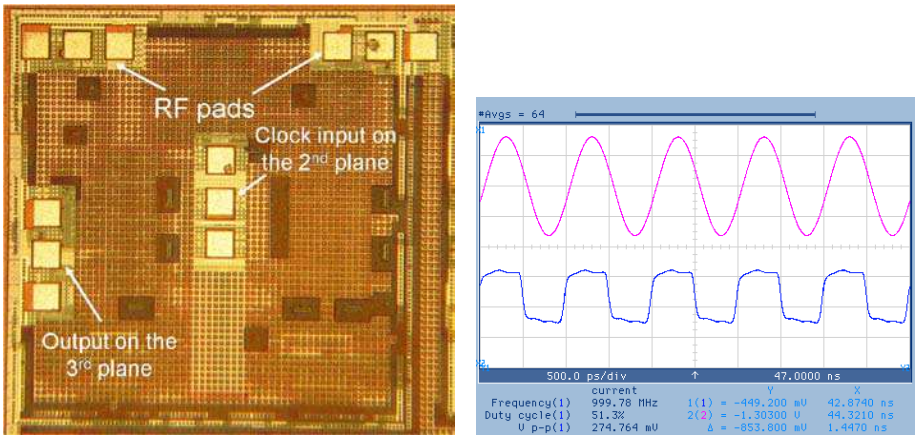


Fig. 9. Fabricated 3-D test circuit. The total area is 3 mm × 3 mm. There are four different blocks, with one input and three output RF pads for each block. The area of each block is approximately 1 mm^2.

The clock distribution networks combine commonly used networks such as H-trees, meshes, and rings. These clock network topologies range from highly symmetric topologies, such as H-trees, as the block shown in Fig. 10a, to fully asymmetric topologies, such as a trunk-based topology. The clock input is a 1.5 V peak-to-peak sinusoidal signal with 0.75 volt DC offset. The clock driver is implemented with a traditional chain of tapered buffers [56], which produces a square waveform at the root of the clock distribution network. The clock distribution network of the block illustrated in Fig. 10a contains a four level H-tree (*i.e.,* equivalent to 16 leaves) with identical interconnect characteristics in each plane. All of the H-trees are connected through a group of interplane vias. Note that the H-tree on the second plane is rotated by 90^o with respect to the H-trees on the other two planes. This rotation effectively eliminates inductive coupling between the H-trees. The second plane is front-to-front bonded with the first plane and both of the H-trees are implemented on the third metal layer. The vertical distance between these clock networks is approximately 2 μm. All of the H-trees are shielded with two parallel lines connected to ground. The waveform shown in Fig. 10b is the clock signal at a leaf of the H-tree on the third plane, demonstrating operation of the circuit at 1 GHz. Experiments

demonstrate that a clock distribution network that combines an H-tree on the second plane and meshes on the other two planes exhibits moderate skew, within 10% of the clock period, and the lowest power consumption [57], [58]. The superior performance of this topology is due to the symmetry of the H-tree and the balancing characteristic of the meshes.



(a) (b)

Fig. 10. Experimental results of the fabricated 3-D circuit, (a) tested circuit block and (b) clock signal waveform from the H-tree on the third plane operating at 1 GHz.

8. Communication Centric 3-D Architectures

A promising design paradigm to appease foreseen interconnect problems is networks-on-chip (NoC) [59], where information is communicated among circuits within packets in an internet-like fashion. The synergy between these two design paradigms, namely NoC and 3-D ICs, can be exploited to significantly improve the performance and decrease the power consumption of communications limited systems. Several interesting topologies that emerge by incorporating the third dimension in networks-on-chip are discussed in this section.

On-chip networks differ from traditional interconnection networks in that communication among the network elements is implemented through the on-chip routing layers rather than the metal tracks of the package or printed circuit board. NoC provide communication among a variety of processing elements (PE), such as processor and DSP cores, memory blocks, FPGAs, and dedicated hardware [60], [61]. Furthermore, the length of the communication channel is primarily determined by the area of the PE, which is typically unaffected by the network structure. Mesh structures have been a popular network topology for conventional 2-D NoC, as illustrated in Fig. 11a, where each processing element (PE) is connected to the network through a router [59].

Integration in the third dimension introduces a variety of topological choices for NoCs. For a 3-D NoC, as shown in Fig. 11b, the total number of nodes is $N = n_1 \times n_2$

$\times n_3$, where n_1, n_2, and n_3 is the number of network nodes in the *x, y,* and *z* direction, respectively. In this topology, each PE is on a single, yet possibly different physical plane (2-D IC – 3-D NoC). In other words, a PE can be implemented on only one of the n_3 physical planes of the system and, therefore, the 3-D system contains $n_1 \times n_2$ PEs on each of the n_3 physical planes, where the total number of nodes is N [62]. A 3-D topology is illustrated in Fig. 11c, where the interconnect network is contained within one physical plane (*i.e.,* $n_3 = 1$), while each PE is integrated on multiple planes, notated as n_p (3-D IC – 2-D NoC). Finally, a hybrid 3-D NoC based on the two previous topologies is depicted in Fig. 11d. In this NoC topology, both the interconnect network and the PEs can span more than one physical plane of the stack (3-D IC – 3-D NoC).

Analytic models of the zero-load latency and power consumption with delay constraints of these networks capturing the effects of the topology on the performance of 3-D NoC have been developed. The overall zero-load network latency for a 3-D NoC is [63]

$$T_{network} = hops(t_a + t_s) + hops_{2-D}t_h + hops_{3-D}t_v + \frac{L_p}{w_c}t_h, \quad (4)$$

where t_a, t_s, t_v, and t_h are the delay of the arbiter, crossbar switch, and vertical and horizontal channels, respectively. *hops*, $hops_{2\text{-}D}$, and $hops_{3\text{-}D}$ denote the average number of hops within the two dimensions n_1 and n_2, and within the third dimension n_3, respectively (see Fig. 11). L_p and w_c denote, respectively, the length of a data packet and the width of the interconnect buss connecting adjacent network routers. L_v denotes the length of the vertical buss, which is equal to one or more TSV lengths.

These models do not incorporate the effects of the routing scheme and traffic load. Since minimum distance paths and no contention are implicitly assumed in these expressions, non-minimal path routing schemes and heavy traffic loads will increase both the latency and power consumption of the network. These models can therefore be treated as lower bounds for both the latency and the power consumption of the network. Alternatively, these expressions provide the maximum improvement in the performance of a conventional NoC that can be achieved with vertical integration.

The resulting decrease in network latency as compared to a standard 2-D IC – 2-D NoC is illustrated in Fig. 12a for increasing network size where the area of each PE is denoted by A_{PE}. The 2-D IC – 3-D NoC topology decreases the number of hops while the interconnect buss delay remains constant. With a 3-D IC – 2-D NoC, the buss delay is smaller but the number of hops remains unchanged. With a 3-D IC – 3-D NoC, all of the latency components can be decreased by assigning a portion of the available physical planes to the network while the remaining planes of the stack are used for the PEs. A decrease in latency of 31.5% and 29.7% can be observed for N = 128 and N = 256 nodes, respectively, with A_{PE} = 1 mm^2. Note that the 3-D IC – 3-D NoC topology achieves the greatest savings in latency by optimally balancing n_3 with n_p. Consequently, the tradeoff between the number of hops and the buss length for various 3-D topologies can be exploited to improve the performance of a network-on-chip.

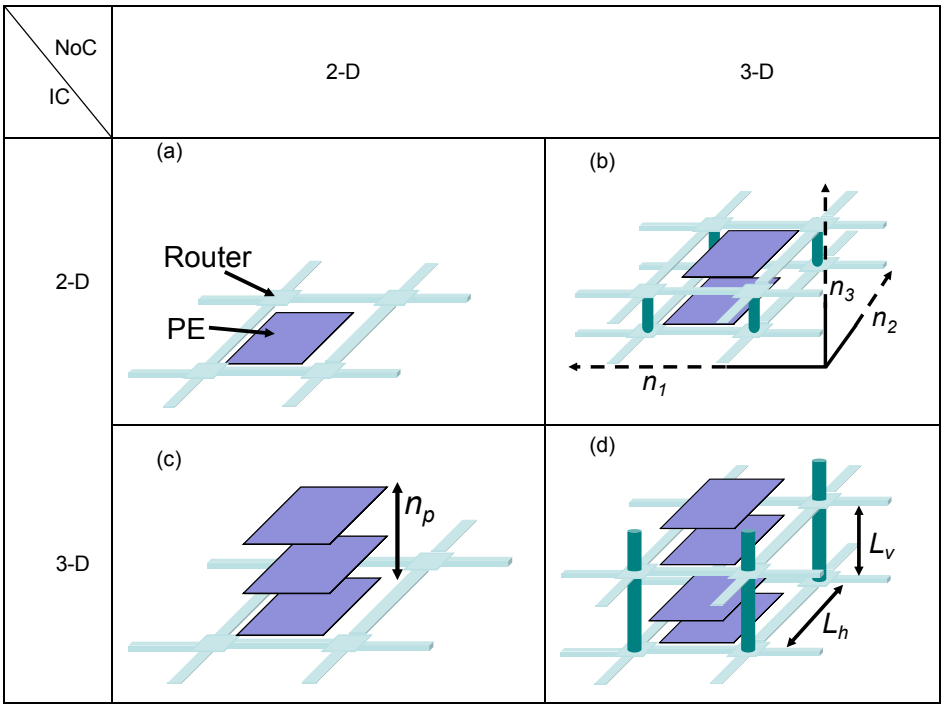


Fig. 11. Various NoC topologies (not to scale), (a) 2-D IC – 2-D NoC, (b) 2-D IC – 3-D NoC, (c) 3-D IC – 2-D NoC, and (d) 3-D IC – 3-D NoC.

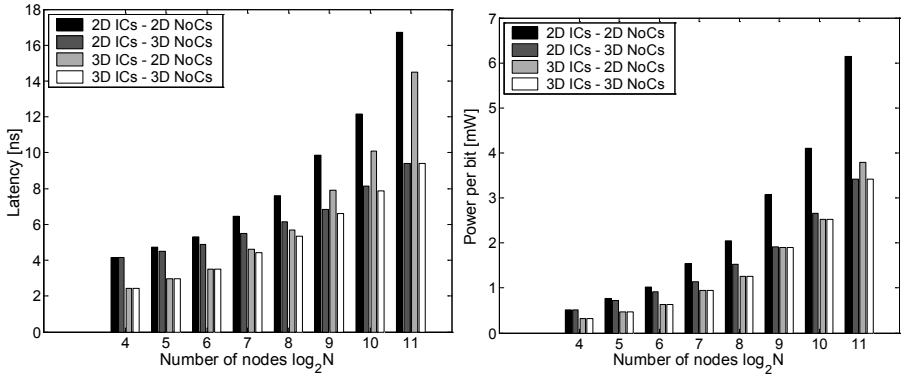


Fig. 12. Performance of 3-D NoC topologies for a range of network sizes where A_{PE} = 1 mm^2; (a) zero-load latency and (b) power consumption with delay constraints.

As with the zero-load latency, each topology affects the power consumption of a network in a different way. The power consumption can be reduced by either decreasing the number of hops that a packet travels or by decreasing the buss length. In Fig. 12b, the power consumption of a 2-D NoC topology is compared to the three-

dimensional topologies previously discussed. A power savings of 38.4% is achieved for $N = 128$ with $A_{PE} = 1$ mm^2. Allowing the available physical planes to be utilized either for the third dimension of the network or for the PEs, the 3-D IC - 3-D NoC scheme achieves the greatest savings in power in addition to the minimum delay.

Note that these topologies emphasize the latency and power consumption of a network, neglecting the performance requirements of the individual PEs. If the performance of the individual PEs is important, only one 3-D topology may be available; however, despite this constraint, a significant savings in latency and power can be achieved since in almost every case the network latency and power consumption are lower than for the 2-D IC – 2-D NoC topology.

9. Conclusions

Developing a design flow for 3-D ICs is a complicated task with many ramifications. Design methodologies at the front end and mature manufacturing processes at the back end are required to effectively provide large scale 3-D systems. A variety of floorplanning, placement, and routing techniques and algorithms for 3-D circuits have been described that consider the unique characteristics of 3-D circuits. In these techniques, the discrete nature of the third dimension is exploited to decrease the number of candidate solutions and, consequently, the computational time required to design a 3-D circuit.

In addition, due to increased power densities and greater distances between the circuits on the upper planes and the heat sink, physical design techniques that embody a thermal objective can be a useful mechanism to manage thermal issues in 3-D ICs. Design techniques can reduce thermal gradients and temperatures in 3-D circuits by redistributing the blocks among and within the planes of a 3-D circuit. Alternatively, thermal vias can be utilized in 3-D circuits to convey heat to the heat sink.

Significant performance improvements can be achieved by optimally placing interplane vias in 3-D circuits. Algorithms for determining the minimum delay of the interplane interconnects are an integral element of the physical design process for 3-D circuits. Interplane interconnect impedances of 3-D circuits vary considerably from 2-D interconnect impedances. This difference is due to several reasons, such as the heterogeneity of 3-D circuits, diverse fabrication technologies, and the variety of bonding styles.

Another requirement for maximizing the speed of 3-D circuits is to reliably distribute the clock signal. A 3-D clock distribution network, however, cannot be directly extended from a 2-D circuit due to the asymmetry of a multi-plane 3-D circuit and the effect of the interplane via impedances. Several clock distribution networks have been developed to investigate synchronization issues in 3-D systems. These clock distribution networks within a three plane 3-D circuit exhibit low clock skew while operating into the gigahertz regime.

In addition to higher performance, 3-D integration offers significant opportunities for designing highly diverse and complex systems. On-chip networks can be a useful solution to provide sufficient communication throughput among the components of these 3-D systems. 3-D NoC are a natural evolution of 2-D NoC, exhibiting superior

performance. These topologies decrease the latency and power consumption by reducing both the number of hops per packet and the length of the communications channels. These 3-D topologies demonstrate the tradeoff between the number of planes required to implement a network and those planes required to implement the PEs. Consequently and not surprisingly, the 3-D IC – 3-D NoC topology achieves the greatest improvement in latency and power consumption by most effectively exploiting the third dimension.

References

[1] "International Technology Roadmap for Semiconductors (ITRS)," Semiconductor Industry Association, 2007.

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

[3] J. W. Joyner *et al.,* "Impact of Three-Dimensional Architectures on Interconnects in Gigascale Integration," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 9, No. 6, pp. 922-928, December 2001.

[4] A. Rahman and R. Reif, "System Level Performance Evaluation of Three-Dimensional Integrated Circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 8, No. 6, pp. 671-678, December 2000.

[5] D. Stroobandt and J. Van Campenhout, "Accurate Interconnection Lengths in Three-Dimensional Computer Systems," *IEICE Transactions on Information and System, Special Issue on Physical Design in Deep Submicron*, Vol. 10, No. 1, pp. 99-105, April 2000.

[6] J. W. Joyner and J. D. Meindl, "Opportunities for Reduced Power Distribution Using Three-Dimensional Integration," *Proceedings of the IEEE International Interconnect Technology Conference*, pp. 148-150, June 2002.

[7] K. Banerjee, S. K. Souri, P. Kapour, and K. C. Saraswat, "3-D ICs: A Novel Chip Design Paradigm for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 602-633, May 2001.

[8] M. Koyanagi *et al.,* "Future System-on-Silicon LSI Chips," *IEEE Micro,* Vol. 18, No. 4, pp. 17-22, July/August 1998.

[9] V. F. Pavlidis and E. G. Friedman, "Interconnect-Based Design Methodologies for Three-Dimensional Integrated Circuits," *Proceedings of the IEEE, Special Issue on 3-D Integration Technology,* Vol. 97, No. 1, pp. 123-140, January 2009.

[10] K. Bernstein *et al.,* "Interconnects in the Third Dimension: Design Challenges for 3-D ICs," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 562-567, June 2007.

[11] R. S. Patti, "Three-Dimensional Integrated Circuits and the Future of System-on-Chip Designs," *Proceedings of the IEEE*, Vol. 94, No. 6, pp. 1214-1224, June 2006.

[12] D. L. Lewis and H.-H. S. Lee, "A Scan-Island Based Design Enabling Pre-Bond Testability in Die-Stacked Microprocessors," *Proceedings of the IEEE International Test Conference,* pp. 1-8, October 2007.

[13] J. Cong, J. Wei, and Y. Zhang, "A Thermal-Driven Floorplanning Algorithm for 3-D ICs," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 306-313, November 2004.

[14] D. Henry *et al.,* "Low Electrical Resistance Silicon Through Vias: Technology and Characterization," *Proceedings of the IEEE International Electronic Components and Technology Conference*, pp. 1360-1365, June 2006.

[15] P. Garrou, "Future ICs Go Vertical," *Semiconductor International* [online], February 2005.

[16] M. Karnezos, "3-D Packaging: Where All Technologies Come Together," *Proceedings of IEEE/SEMI International Electronics Manufacturing Technology Symposium,* pp. 64-67, July 2004.

[17] J. Miettinen, M. Mantysalo, K. Kaija, and E. O. Ristolainen, "System Design Issues for 3D System-in-Package (SiP)," *Proceedings of the IEEE International Electronic Components and Technology Conference*, pp. 610-615, June 2004.

[18] E. Beyne, "The Rise of the 3rd Dimension for System Integration," *Proceedings of the IEEE International Interconnect Technology Conference,* pp. 1-5, June 2006.

[19] V. F. Pavlidis and E. G. Friedman, "Interconnect Delay Minimization through Interlayer Via Placement in 3-D ICs," *Proceedings of the ACM Great Lakes Symposium on VLSI,* pp. 20-25, April 2005.

[20] *"FDSOI Design Guide,"* MIT Lincoln Laboratories, Cambridge, 2006.

[21] J. A. Burns *et al.,* "A Wafer-Scale 3-D Circuit Integration Technology," *IEEE Transactions on Electron Devices,* Vol. 53, No. 10, pp. 2507-2515, October 2006.

[22] C. A. Bower *et al.,* "High Density Vertical Interconnect for 3-D Integration of Silicon Integrated Circuits," *Proceedings of the IEEE International Electronic Components and Technology Conference*, pp. 399-403, June 2006.

[23] D. M. Jang *et al.,* "Development and Evaluation of 3-D SiP with Vertically Interconnected Through Silicon Vias (TSV)," *Proceedings of the IEEE International Electronic Components and Technology Conference*, pp. 847-850, June 2007.

[24] I. Savidis and E. G. Friedman, "Electrical Modeling and Characterization of 3-D Vias," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 784-787, May 2008.

[25] R. H. J. M. Otten, "Automatic Floorplan Design," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 261-267, June 1982.

[26] E. F. Y. Yong, C. C. N. Chu, and C. S. Zion, "Twin Binary Sequences: A Non-Redundant Representation for General Non-Slicing Floorplan," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 22, No. 4, pp. 457-469, April 2003.

[27] L. Cheng, L. Deng, and D. F. Wong, "Floorplanning for 3-D VLSI Design," *Proceedings of the IEEE International Asia and South Pacific Design Automation Conference*, pp. 405-411, January 2005.

[28] Z. Li *et al.,* "Hierarchical 3-D Floorplanning Algorithm for Wirelength Optimization," *IEEE Transactions on Circuits and Systems I: Regular Papers,* Vol. 53, No. 12, pp. 2637-2646, December 2006.

[29] Y. Deng and W. P. Maly, "Interconnect Characteristics of 2.5-D System Integration Scheme," *Proceedings of the IEEE International Symposium on Physical Design*, pp. 341-345, April 2001.

[30] Z. Li *et al.,* "Efficient Thermal Via Planning Approach and its Application in 3-D Floorplanning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 26, No. 4, pp. 645-658, April 2007.

[31] T. Yan, Q. Dong, Y. Takashima, and Y. Kajitani, "How Does Partitioning Matter for 3D Floorplanning," *Proceedings of the ACM International Great Lakes Symposium on VLSI,* pp. 73-76, April/May 2006.

[32] [Online] Available: http://www.cse.ucsc.edu/research/surf/GSRC/progress.html

[33] M. Healy *et al.,* "Multiobjective Microarchitectural Floorplanning for 2-D and 3-D ICs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 26, No. 1, pp. 38-52, January 2007.

[34] W.-C. Lo *et al.,* "An Innovative Chip-to-Wafer and Wafer-to-Wafer Stacking," *Proceedings of the IEEE International Electronic Components and Technology Conference*, pp. 409-414, June 2006.

[35] B. Goplen and S. Sapatnekar "Placement of Thermal Vias in 3-D ICs Using Various Thermal Objectives," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 25, No. 4, pp. 692-709, April 2006.

[36] S. T. Obenaus and T. H. Szymanski, "Gravity: Fast Placement for 3-D VLSI," *ACM Transactions on Design Automation of Electronic Systems,* Vol. 8, No. 3, pp. 298-315, July 2003.

[37] A. Harter, *Three-Dimensional Integrated Circuit Layout,* Cambridge University Press, 1991.

[38] I. Kaya, M. Olbrich, and E. Barke, "3-D Placement Considering Vertical Interconnects," *Proceedings of the IEEE International SOC Conference*, pp. 257-258, September 2003

[39] W. R. Davis *et al.,* "Demystifying 3D ICs: the Pros and Cons of Going Vertical," *IEEE Design and Test of Computers,* Vol. 22, No. 6, November/December 2005.

[40] H. Eisenmann and F. M. Johannnes, "Generic Global Placement and Floorplanning," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 269-274, June 1998.

[41] B. Goplen and S. Sapatnekar, "Efficient Thermal Placement of Standard Cells in 3-D ICs using a Force Directed Approach," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 86-89, November 2003.

[42] [Online] Available: http://er.cs.ucla.edu/benchmarks/ibm-place

[43] [Online] Available: http://www.cbl.ncsu.edu/pub/Benchmark_dirs/LayoutSynth92

[44] R. J. Enbody, G. Lynn, and K. H. Tan, "Routing the 3-D Chip," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 132-137, June 1991.

[45] C. C. Tong and C.-L. Wu, "Routing in a Three-Dimensional Chip," *IEEE Transactions on Computers,* Vol. 44, No. 1, pp. 106-117, January 1995.

[46] J. Minz and S. K. Lim, "Block-Level 3-D Global Routing With an Application to 3-D Packaging," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 25, No. 10, pp. 2248-2257, October 2006.

[47] T. Ohtsuki, *Advances in CAD for VLSI: Vol. 4*, *Layout Design and Verification*, Elsevier, 1986.

[48] J. Cong, M. Xie, and Y. Zhang, "An Enhanced Multilevel Routing System," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 51-58, November 2002.

[49] J. Cong and Y. Zhang, "Thermal Driven Multilevel Routing for 3-D ICs," *Proceedings of the IEEE Asia and South Pacific Design Automation Conference,* pp. 121-126, June 2005.

[50] V. F. Pavlidis and E. G. Friedman, "Timing Driven Via Placement Heuristics in 3-D ICs," *Integration, the VLSI Journal,*Vol. 41, No. 4, pp. 489-508, July 2008.

[51] K. D. Boese *et al.*, "Fidelity and Near-Optimality of Elmore-Based Routing Constructions," *Proceedings of the IEEE International Conference on Computer Design,* pp. 81-84, October 1993.

[52] J. Löfberg, "YALMIP: A Toolbox for Modeling and Optimization in MATLAB," *Proceedings of the IEEE International Symposium on Computer-Aided Control Systems Design*, pp. 284-289, September 2004.

[53] E. G. Friedman (Ed.), *Clock Distribution Networks in VLSI Circuits and Systems,* IEEE Press, New Jersey, 1995.

[54] E. G. Friedman, "Clock Distribution Networks in Synchronous Digital Integrated Circuits," *Proceedings of the IEEE,* Vol. 89, No. 5, pp. 665-692, May 2001.

[55] T. Xanthopoulos *et al.,* "The Design and Analysis of the Clock Distribution Network for a 1.2 GHz Alpha Microprocessor," *Proceedings of the IEEE International Solid-State Circuits Conference,* pp. 402-402, February 2001.

[56] C. Punty and Laszlo Gal, "Optimum Tapered Buffer," *IEEE Journal of Solid-State Circuits*, Vol. 27, No. 1, pp. 1005-1008, January 1992.

[57] V. F. Pavlidis, I. Savidis, and E. G. Friedman, "Clock Distribution Networks for 3-D Integrated Circuits," *Proceedings of the IEEE International Conference on Custom Integrated Circuits,* pp. 651-654, September 2008.

[58] V. F. Pavlidis, I. Savidis, and E. G. Friedman, "Clock Distribution Architectures for 3-D SOI Integrated Circuits," *Proceedings of the IEEE International Silicon-on-Insulator Conference,* pp. 111-112, October 2008.

[59] A. Jantsch and H. Tenhunen, *Networks on Chip*, Kluwer Academic Publishers, 2003.

[60] L. Benini and G. De Micheli, "Networks on Chip: A New SoC Paradigm," *IEEE Computer*, Vol. 31, No. 1, pp. 70-78, January 2002.

[61] S. Kumar *et al.,* "A Network on Chip Architecture and Design Methodology," *Proceedings of the International IEEE Annual Symposium on VLSI*, pp. 105-112, April 2002.

[62] F. Li *et al.,* "Design and Management of 3D Chip Multiprocessors Using Network-in-Memory," *Proceedings of the IEEE International Symposium on Computer Architecture*, pp. 130-142, June 2006.

[63] V. F. Pavlidis and E. G. Friedman, "3-D Topologies for Networks-on-Chip," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* Vol. 15, No. 10, pp. 1081-1090, October 2007.