Designing a 3D tree-based FPGA: Optimization of butterfly programmable interconnect topology using 3D technology

Vinod Pangracious, Habib Mehrez, Zied Marrakchi

Designing a 3D Tree-based FPGA: Optimization of Butterfly Programmable Interconnect Topology Using 3D Technology

Vinod Pangracious, and Habib Mehrez
LIP6, University of Pierre and Marie Curie Paris VI, France 75006
Email: Vinod.Pangracious@etu.upmc.fr

Zied Marakchi
FLEXRAS Technologies Paris, France
Email: Zied.Marrakchi@flexras.com

***Abstract*—The CMOS technology scaling has greatly improved the overall performance and density of the Mesh-based Field Programmable Gate Arrays (FPGAs), nonetheless the gap between FPGAs and ASICs in terms of logic density, speed and power consumption remains very wide mainly due the programming overhead. The logic density and area overhead is improved by using Tree-based FPGA architecture using Butterfly-Fat-Tree (BFT) based network topology. However the wire-length increases exponentially as the tree grows to higher levels. We have introduced a horizontally partitioned 3-dimensional (3D) design methodology to optimize the BFT based programmable interconnect delay of the Tree-based FPGA. In this paper we describe a 2 tier horizontally partitioned 3D stacked Tree-based FPGA demonstrator, designed and implemented using Tezzaron's 130nm, 3D technology. We finally evaluate the speed and area overhead of the proposed 3D Tree-based FPGA using the newly developed experimental design and evaluation methodology and show that the horizontally partitioned BFT programmable interconnect topology based 3D Tree-based FPGA improves speed by 2.06 times and reduce interconnect area by 2.8 times compared to 3D Mesh-based FPGA with identical logic resources.**

I. Introduction

Modern Field Programmable Gate Arrays (FPGAs) have become a viable alternative to cell based design technology by providing re-configurable computing platforms with improved performance and density. With the onset of sub-100nm CMOS technologies, the design and prototyping cost of the cell-based custom integrated circuit implementation have become exorbitant for most ASICs, making FPGAs increasingly popular. With their regular structure, they also scale easily with sub-100nm technologies. Current FPGAs, however cannot meet the area and speed requirements of many ASICs due to their high programming overhead. Recent studies shows that in Mesh-based industrial FPGAs, 80% of overall design delay and 90% of the chip area are attributed to programmable routing resources [1], [2] [3]. It has also reported that in Mesh-based FPGA, programmable interconnects contributes as much as 60% of the total dynamic power consumption [4]. Considering the area, delay and power consumption, the programmable interconnects are the key elements in FPGA design [2], [5].

The 3D integrated circuit (IC) technology has emerged as one of the most promising solutions for overcoming the challenges in interconnection and integration complexity in modern circuit designs [6]. The 3D integration technology can effectively reduce global interconnect length and increase circuit performance without increasing the power consumption. Through Silicon Vias (TSVs) are the key enabling technology element for 3D integration, which is currently being actively evaluated as a potential solution to reduce the interconnect delay and increase the logic density in FPGA. Recently, multiple technology and product demonstrations of TSV in silicon interposer have been reported for high performance FPGA applications [18], [19], [20]. Based on the design and manufacturing specifications and maturity of enabling technologies, a pattern in technology adoption is beginning to emerge. Three-dimensional integration where TSVs are incorporated in active device layers, is considered to be the Holy-Grail of vertical die stacking. However the recent product demonstrations from major FPGA manufactures revels the adoption of silicon interposer based 3D integration scheme, where TSVs are incorporated in passive silicon interposer [19], [20]. The primary focus of this paper is to demonstrate a 3D integration scheme to partition and optimize the multilevel programmable interconnect network of Tree-based FPGA based on BFT network topology, where TSVs are incorporated in active layers of the 3D chip.

There are two major types of 3D FPGA architectures found in the literature. The first one is developed by monolithic stacking, whereby the active devices are lithographically built in between metal layers [21] and the second type is evolved from original 2D structure by extending the 2D switch boxes (SBs) to 3D ones [11], [12]. So far, there are two design and exploration frameworks targeting 3D FPGA architectures: the three-dimensional place and route (TPR) [12] and 3D MEANDER [11]. In TPR, all SBs are assumed to be 3D-SBs and the number of TSVs is assumed to be unlimited, which is an impractical assumption as far as design and manufacturing of 3D chips is concerned. Meanwhile 3D MEANDER is a fully-fledged design framework for 3D FPGAs and it provides the capability to analyze the impact of different deployment strategy for 3D-SBs in multi-tier FPGAs. It proposes a family of 3D FPGA architectures in which 2D-SBs and 3D-SBs are intermittently used in certain regular spatial patterns. Nonetheless the number of available TSVs within 3D-SBs is assumed to be fixed and that means the design does not investigate the impact of different numbers of TSVs in a 3D-SB.

II. Motivation And Problem Formulation

According to [11], [12] the SBs has been the most area-consuming unite compared to other design elements in 2D FPGAs and this situation is becoming even worse in 3D

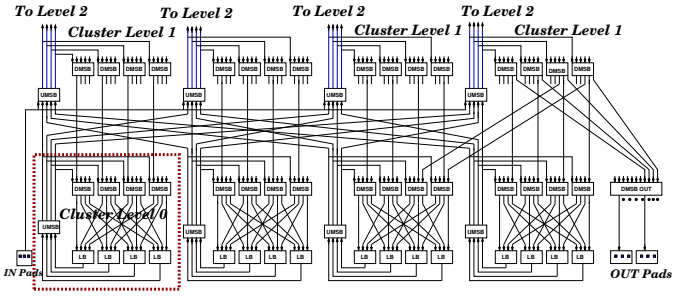

Fig. 1. A 2 level Tree-based Multilevel FPGA interconnect: Upward and Downward Interconnection network

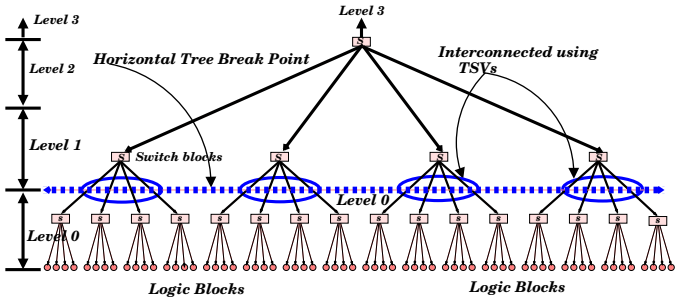

Fig. 2. Tree-based FPGA: symbolic representation of horizontal break-point

FPGAs because the TSVs are located in 3D-SBs. Although the design and manufacturing engineers are trying to reduce TSV dimensions, the minimum feature size on the die is also shrinking. Therefore, the TSVs are expected to remain larger than wire dimensions in metal layers within the die [9]. Moreover it has been reported in [25] that the TSV utilization is actually quite low if the 3D-SBs are with full vertical connectivity in use. The experiments carried out in our laboratory and recent publications points out that the utilization of TSVs is actually very low in 3D Mesh-based FPGAs with full vertical connectivity [25], which motivates us to explore new architectures that can be better optimized to achieve higher speed, logic density, reduced power consumption and area to minimize the gap between FPGAs and ASICs. In this paper, we prefer to use a Tree-based multilevel FPGA architecture, because from our experimental and design experience, we believe, due to the BFT based multilevel interconnect topology, Tree-based FPGA architectures is more suitable to build high density 3D re-configurable systems compared to Mesh-based industrial FPGAs [8]. In a Tree-based FPGA architecture [8], [10], the programmable interconnects are arranged in a BFT based multilevel network with the switch blocks placed at different tree levels and the logic blocks (LBs) are grouped into clusters. Due to the multilevel network arrangement, we do not have to deal with 3D SBs in the case of Tree-based FPGA, rather all the switch blocks remain as 2D and only the signal and I/O communications that are partitioned between multi-tiers are interconnected using TSVs.

In a Tree-based FPGA architecture [10], The LBs are grouped into clusters and each cluster contains a switch block to connect local LBs. A switch block is divided into Mini switch Blocks (MSBs). The Tree-based FPGA architecture unifies two unidirectional upward and downward interconnection networks by using a *Butterfly-Fat-Tree* (BFT) topology to connect the Downward MSBs (DMSBs) and the Upward MSBs (UMSBs) to LBs inputs and outputs. As illustrated in Figure 1, the UMSBs are used to allow LBs outputs to reach a large number of DMSBs and to reduce fanout on feedback lines. The UMSBs are organized in a way allowing LBs belonging to the same *owner-cluster* to reach exactly the same set of DMSBs at each level. Thus positions, inside the same cluster, are equivalent, and LBs can negotiate with their siblings about the use of a larger number of DMSBs depending on their fanout. Therefore the external IO pads, clusters or logic-block positions inside the direct owner cluster become equivalent and there is no need to re-arrange them anymore. The input and output pads are grouped into specific clusters and are connected to UMSBs and DMSBs, respectively as presented in Figure 1. Thus, all input and output pads can reach any LBs of the architecture. The programmable interconnects of a Tree-based FPGA architecture are arranged in a multilevel network with the switch blocks placed at different tree levels using a *Butterfly-Fat-Tree* network topology. A 3D multilevel interconnect network organization using horizontally partitioned design methodology with the *horizontal break-point* placed between Tree level 3 and 4 is presented in [8]. The placement of *break-point* is based on the interconnect delay measurement results of individual Tree levels. Figure 2 illustrates the horizontal partitioning methodology of 2 level, arity 4 (cluster with LBs) Tree-based FPGA. The horizontal partitioning in done in such a way to stack the BFT based programming overhead of Tree-based FPGA on top of LBs and interconnected using TSVs.

III. Summary of Results and Outline of Paper

According to the report [24], the Butterfly-Fat-Tree based FPGA interconnect can be placed in *O(N)* area using *O(log(N))* metal layer in 2-dimensional area. In our 3D physical design methodology, we introduced a horizontal *break-point* to partition the multilevel Butterfly-Fat-Tree based programmable interconnect network into two independent designs interconnected using TSVs. Figure 2 illustrate the horizontal partitioning methodology of a 2 tier 3D Tree-based FPGA. In the horizontally partitioned 3D design process, the logic blocks along with local programmable interconnects are placed in tier 1 and programmable interconnects belong to tree levels above the horizontal break point are placed in tier 0. Using our comprehensive experimental setup, we investigated the performance metrics and we show that 3D stacked Tree-based FPGA can achieve 4.2 times higher logic density, 2.1 times lower critical path delay and 1.8 times lower power consumption than the baseline 3D Mesh-based industrial FPGA. The next section IV presents the experimental software flow and 3D design methodology developed for this experiment. In section V, we describe the horizontal partitioning methodology, timing analysis and 3D design flow used to design our 2 tier 3D Tree-based FPGA demonstrator. Section VI describe the 3D physical design and stacking methodology of Tree-based FPGA. In section VII we explain the experimental methodology developed to quantify the improvement in speed, architecture area and design optimization. In section VIII, we quantify the reduction in TSV count, programmable routing resources and area. In section IX we describe, how to balance and remove inter-layer heat from 3D FPGA chip using spatial distribution of TSVs and power delivery networks (PDNs) controlled 3D thermal model and design optimization.

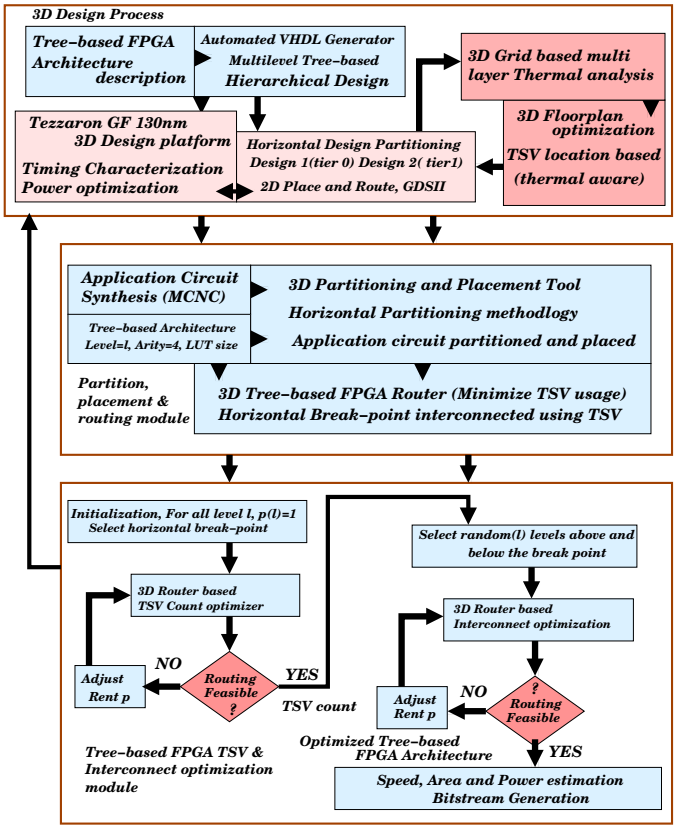

Fig. 3. 3D Tree-based FPGA design and experimentation software Flow

IV. Design and Implementation Methodology

The proposed software flow for design and experimentation of 3D Tree-based FPGA architecture is illustrated in Figure 3. The HDL code generator is designed to generate VHDL code based on a hierarchical design approach that partitions the design into smaller sections, which implement clusters separately and assemble them together at the final design phase. The physical design experiments were performed using the layout generated with the help of Global Foundries 130nm technology node (Tezzaron 3D Design platform). Mentor's spice accurate circuit simulator *Eldo* is used to estimate the wire delay and power consumption of switches and interconnection networks at different tree levels of 2D layout [8]. The design and development strategy of 2D physical design and 3D floorplan development of Tree-based FPGA is presented in [13]. The floorplan tool is augmented with flexibility to partition the Tree-based BFT interconnect network horizontally based on the design specifications of the 3D stacked Tree-based FPGA.

Thermal issues in FPGAs have been relatively unexplored for a long time. Few reports [14], [15] have proposed the use of distributed sensors for monitoring temperatures in FPGAs. However, they consider only LBs in the FPGA fabric, and consequently, observed very little temperature variations across the die. Nevertheless the programmable interconnect network of FPGA consumes a lot of power as well. In contrast, the methodology we used, considers the spatial distribution of signal TSVs, power delivery networks (PDNs), power consumption of computational and interconnect sections separately to investigate the temperature variations in FPGA. First we measure the 2D block level temperature estimation and later on the temperature variation of the 2 tier 3D Tree-based FPGA system with proposed TSV based thermal optimization solutions are evaluated. The thermal model presented in [16] is augmented to extract the thermal profile of the multi-layer chip based on geometrical features and power consumption of the FPGA functional units while considering the spatial distribution of TSVs and PDNs of the layout.

The design netlist to be mapped is obtained in .NET format. The LUTs and I/Os are first partitioned into different clusters in such a way that the inter-cluster communication is minimized while considering the horizontally partitioned programmable interconnects. After completing the partitioning, placement file is generated. It contains positions of different blocks in the architecture. This placement file along with netlist file and Tree-based FPGA architecture specification files are then passed to 3D router, which is responsible for routing the application netlist. The routing problem consists of assigning nets that interconnects the already placed LBs at tier 1 and routing resources in the multilevel programmable interconnect structure placed at tier 0 of the 3D Tree-based FPGA. The router algorithm is based on *PathFinder* [10], that uses an iterative, negotiation-based approach to successfully route all nets in application netlist. The 3D timing analyzer generates direct acyclic timing graph of the routed circuit of the multi-tier 3D chip to evaluate the critical path delay of 3D stacked Tree-based FPGA. Based on routing result, the different sub-paths are identified and each edge is annotated with delay of corresponding sub-path. The edges that interconnect different tiers of the 3D Tree-based FPGA annotate corresponding TSV delay to the pins which the circuit specifies as connected between tier 0 and 1 (3D nets). In order to optimize the TSV count and routing resources, a Rent-based wire length distribution model implemented using 3D router is used. After completing the TSV count and architecture optimization, the router will estimate the critical path delay, TSV count, area and power consumption of the optimized 2 tier 3D FPGAs.

V. 3D Design Methodology

In horizontal partitioning methodology, the location of the break point is decided based on optimization of interconnect network delay. The interconnect delay of Tree-based architecture increases exponentially [8], [10] as the Tree grows to higher levels. In case of horizontal partitioning methodology, the LBs and local interconnect levels below the break-point are placed on tier 1 and programmable interconnect resources along with I/Os at levels above break point are placed in tier 0 of the 3D stacked chip. Figure 4 shows the 2 tier 3D layout representation of the Tree-based FPGA along with I/Os connected to level 4, 5 and 6 and the break-point is set between level 3 and 4. The location of break-point is decided based on the delay measurements of tree levels illustrated in Figure 6 where the delay between level 3 to 4 is greater than 2ns. We have approximately 20% white space in tier 0 of horizontally partitioned homogeneous Tree-based FPGA due to unbalanced hardware partition. As illustrated in Figure 5, the partition is done in such a way to stack programmable interconnects overhead of FPGA on top of logic blocks and this method provides more flexibility in increasing logic density and lowering critical path delay. As an extension of this project, additional Hard-Blocks (HBs) were integrated into

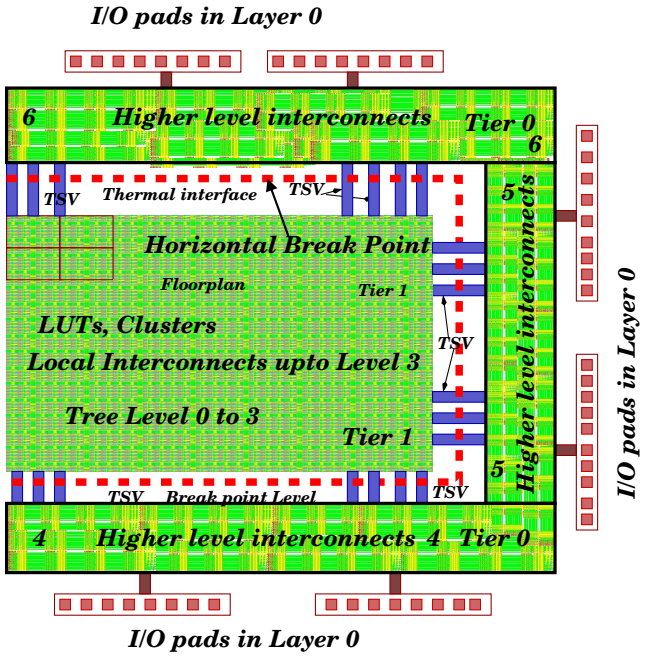

Fig. 4. Horizontal partitioning: break point between *level 3 and 4* of the 3D Tree-based FPGA

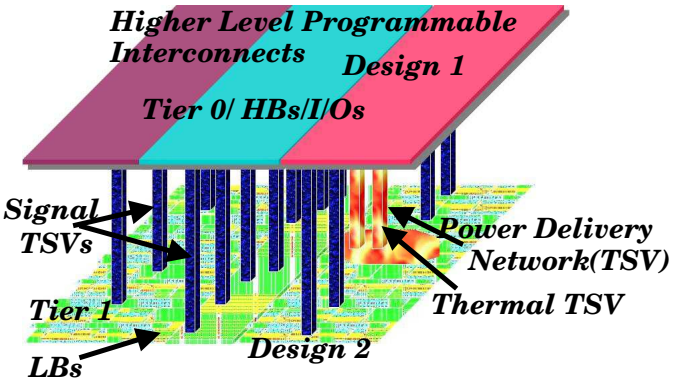

Fig. 5. 2 Tier 3D Tree-based FPGA structure with programmable overhead on top of LBs

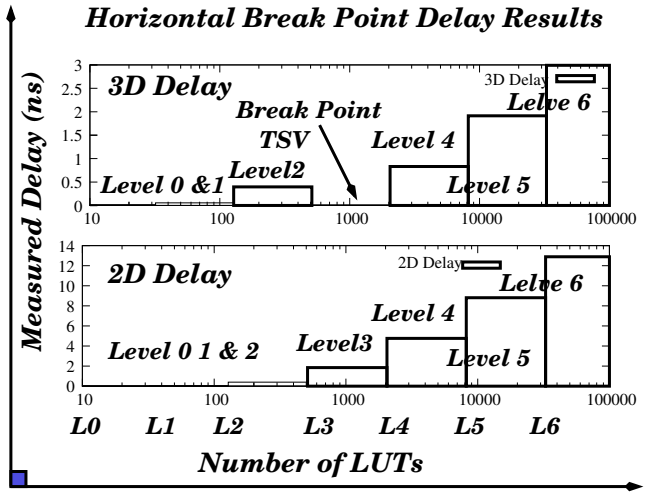

Fig. 6. Horizontal break-point interconnect delay estimation of 7 level Tree-based FPGA architecture

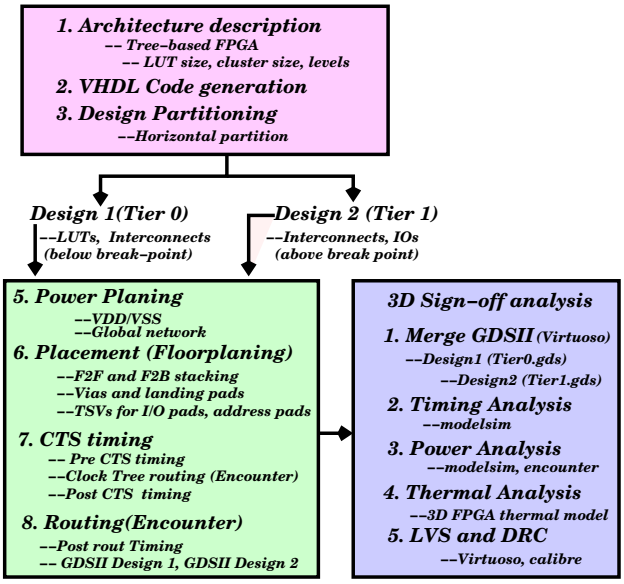

Fig. 7. 3D design flow for design and implementation of horizontal partitioned 3D Tree-based FPGA

the available white space in tier 0 to design and implement 3D heterogeneous Tree-based FPGA, but heterogeneous FPGA design model is beyond the scope of this paper.

We used 6 metal, 130nm process node provided by Global Foundries that is modified to include TSVs according to the specification of Tezzaron Semiconductor. Tezzaron's via-first 3D manufacturing process produce very small TSVs that are approximately 1.2μm wide with 2.5μm minimum pitch and 6μm height [22]. The area around the TSV has been expanded to include *keep out zones* [22] to make TSVs fit within 5 to 6 standard cell area, since the area of a 4-transistor logic gate is projected to range from $0.82\mu m^2$ to $0.20\mu m^2$ by 2015 [22]. The TSV *keep out zones* are essential for 3D design to maintain the performance of active device placed adjacent to TSVs. The measured values of TSV resistance R_{TSV} and capacitance C_{TSV} are $\approx 600m\Omega$ and 15fF respectively. The wire delay estimation of Tree levels for the 3D stacked Tree-based FPGA is extracted from the 2 tier layout developed using Tezzaron Process [8]. The break-point interconnect (TSVs) delay is optimized using the TSV model presented in [11], [23]. The TSV delay estimated using *eldo* is $\approx 28pS$. In tier 0, the spatial distribution of TSVs and interconnect switches along with I/Os are rearranged in order to optimize the wire delay at higher levels. The measured interconnect delay for 2D and 3D layouts illustrated in Figure 6.

VI. 3D Physical Design Methodology

Figure 7 shows the overall physical design methodology used for the design and implementation 2 tier 3D Tree-based FPGA. The physical design process begins with the RTL description of Tree-based FPGA generated using VHDL code generator as described in section IV. In the case of horizontal partitioning, tier 1 contains LUTs and local programmable

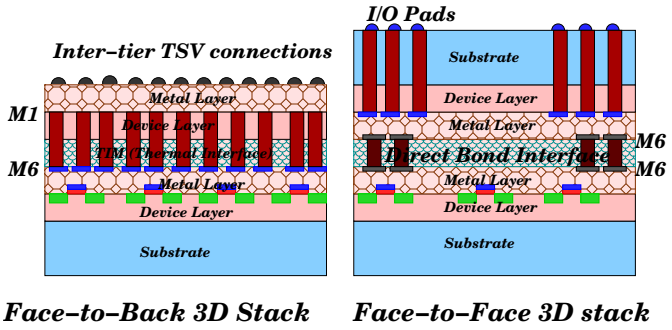

Fig. 8. Side view of the stacked dies based on Tezzaron's F2B and F2F stacking technology

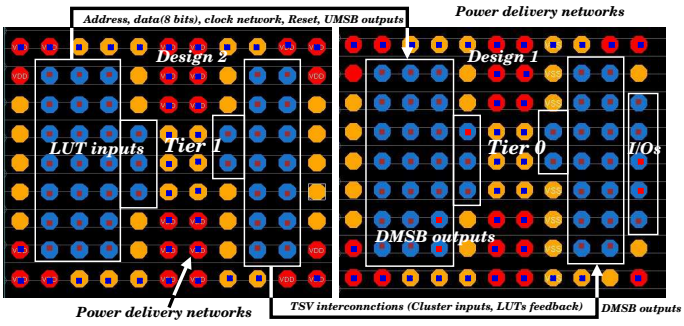

Fig. 9. TSV assignment on Tier 1 (design 2) and Tier 0 (design 1) of the 3D stacked Tree-based FPGA

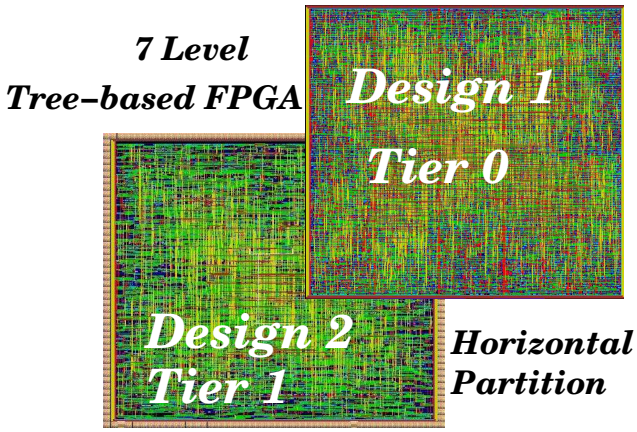

Fig. 10. Tier 1 (design 2) and Tier 0 (design 1) of the horizontally partitioned 7 level Tree-based FPGA with 16K LUTs

interconnects from levels 0 to 3 (design2) and tier 0 contains programmable interconnect above the break-point along with I/Os (design1) as illustrated in Figure 4. We then used Cadence design compiler to compile VHDL into structural Verilog for each die. The compiled Verilog is then input into Cadence Encounter to perform semi-automated physical design steps. The design tool augmented to test different 3D stacking methodologies. We used both Face-to-Face (F2F) and Face-2-Back (F2B) stacking methodology provided by Tezzaron's 3D design platform using via first TSV process. Tezzaron's 3D stacking kit support two types of TSV structures: Supper-Contact and Super-Via. In our design we used Supper-Contact, using tungsten via fill. In F2F stacking, the die is flipped from left-to-right and bonded using 3.4μm Metal 6 pads with 5μm pitch. After bonding the tier 0 die is thinned down to the TSV, that is 6μm. After thinning the tier 0 is about 12μm thick, with about 6μm metal and wiring plus 6μm of bulk silicon with TSVs. The insulation material between TSV and silicon is oxide with 1000 A^o thickness.The I/O signals are routed through TSVs to the back surface of tier 0 and from there, they will be fanned out past the edge of the device to connect to I/O pads on the surface of the 3D FPGA chip, while in F2B stacking, the tier 0 via-first TSVs have their landing pads on Metal 1 and Metal 6. The connection between via-first TSVs are made using local interconnection and vias in between adjacent dies. The tier 0 die is thinned down to TSVs first and bonded using the TSV landing pads. These landing pads include *keep-out-zones* uniformly located around them to reduce coupling effects on active devices located around it.

A. 3D Stacking Methodology: F2F and F2B TSV Placement

Communication between LBs in tier 1 and higher level routing resources in tier 0 established using Face-2-Back (F2B) and Face-2-Face (F2F) vias as presented in Figure 8. Any net that connects to F2B or F2F via and thus interconnecting circuitry from tier 1 to tier 0, is defined as 3D net. The individual designs for each die (tier 0 and 1) therefore must contain pins for all nets that cross the vertical interconnection boundary of the 2 tier test chip. The test chip contains 7 Tree levels and horizontal break-point between levels 3 and 4 and 16K LUTs. The tier 1 (design2) contains only LBs and their communications to tier 0 (design 1). The interconnecting vias from tier 0 and 1 were manually placed on both dies for F2F and F2B stacking. In the case F2F stacking the TSVs are used only for off-chip I/O and power/ground connections. One exclusive requirement that the Tezzaron TSV process imposes is on mandatory minimum TSV pitch of 250μm throughout the entire wafer. This requirement forces to include at least one TSV inside every 250μm window in our design. According the Tezzaron 3D process, this is used for planarity of the wafer during Chemical and Mechanical Polishing (CMP) process. In the case of F2B back stacking the tier 0 die is thin down to TSV, that is 6μm in length before bonding, while in F2F stacking, the die is thinned down to TSV after bonding without handling the wafer or die. In certain locations, we manually inserted dummy TSVs before placement to meet this requirement. Figure 9 present the TSV assignment of tier 0 and 1 dies of a cluster in the 3D stacked Tree-based FPGA test chip. In our 3D Tree-based FPGA test chip, we have 7 Tree levels with arity set to 4. The test chip includes 16K LUTs and 5120 TSVs connections between tier 0 and 1. The power and ground distribution networks (PDNs) are mainly generated using the strip and ring generation commands in cadence Encounter.

B. 3D Placement and Route

Cadence Encounter is used to perform the placement and routing of both tier 0 and tier 1 dies. The 3D inter-tier communication information is propagated to the placer through the use of fixed pins on metal 6 for tier 0 and TSVs for tier 1 representing F2F and F2B connections. These pins impose certain restrictions on placement and optimization of the full 3D chip, but the spatial distribution of TSVs and PDNs were also considered to check the thermal profile of the 3D

stacked chip. We try to balance the design complexity and inter-tier temperature optimization using 3D thermal model. Cadence Encounter is also used to perform sizing and buffering optimization, and NanoRout is used to perform routing. The inter-tier connection information is propagated through to the optimizer and router through back-annotation of capacitance and arrival time requirements on the fixed pins. Figure 10 presents layout views of tier 0 and tier 1 of the horizontally partitioned 7 level 16K LUTs 3D stacked Tree-based FPGA.

C. 3D Sign-Off Analysis

The existing Cadence, Synopsys and Mentor Graphics tools are designed for 2D, thus do not have the capability to handle designs with TSV interconnections. The 3D design timing analysis is based on Synopsys PrimeTime. At first the Verilog netlist files of tier 0 and 1 designs are prepared and the SPEF files containing extracted parasitic values for all the nets of the designs. Then we create the top level Verilog netlist that instantiates the design of each die and connects the 3D nets using Supper-Contact (TSVs) based on F2B connections or direct boding connections for F2F stacking. Our experimental design flow includes 3D Thermal modeling as illustrated in Figure 3. Hotspot power dissipation results in significantly higher temperatures in 3D stacked chips compared to the same power dissipation in single 2D chips. The reason this temperature increase is due to the reduced thermal spreading in the thinned dies on the one hand, and to the use of low thermal conductivity adhesives on the other hand. In our design, the tier 0 die is thin down to TSV, that is $6\mu m$ and another $6\mu m$ metal layers. Therefore a detailed thermal analysis at the design stage is required. For this reason we have integrated a 3D thermal resistance mesh based multi-layer thermal model in design flow. The 3D thermal model consider spatial distribution of TSVs and PDNs while estimating the thermal profile of the 3D chip. In certain case the thermal solution suggest if we need additional dummy TSVs to balance or remove the inter-layer heat. However this is a limited operation, since TSVs consume more area compared to metal wires.

VII. 3D Tree-based FPGA Performance Evaluation

In order to validate the performance of horizontally partitioned 3D Tree-based FPGA architecture, we used a fully connected (Rent=1) Tree-based FPGA architecture with 7 levels and arity 4 for each benchmarks circuits. Once the partitioning is over, the individual netlist are placed and routed using the experimental flow described in section IV excluding the architecture optimization section. The critical path delay analysis of horizontal partitioning methodology is reported in Table I. The average improvement in speed measured for horizontally partitioned stacking methodology is 64.1%. The horizontally partitioned 3D stack methodology performed 2.06 times better compared to 3D Mesh-based Industrial FPGA with identical logic resources [11]. The main reason for speed improvement in horizontal partitioning method is due to shortening of wire length between levels 3 and 4 (horizontal break-point) using TSVs and optimization of the interconnect delay at the higher level Tree networks that are placed in tier 0 of the 3D stacked chip as illustrated in Figure 5. The 3D Mesh-based FPGA reported in [11] with intermittent

TABLE I. 2 tier 3D Tree-based FPGA Performance Analysis

	Tree Levels=7, Arity=4, Arch=4x4x4x4x4x4x4			
	Critical Path delay (nS)		Performance Gain(%)	
Name MCNC	Tree-based 2D (ns)	Horizontal 3D (ns)	2D Vs 3D Horizon Gain (%)	3D Mesh FPGA Gain (%)
alu4	59.91	25.81	56.91	36.1
apex2	80.41	30.92	61.54	50
apex4	76.42	31.83	58.34	48.1
bigkey	79.1	20.19	74.48	41.7
clma	198.6	59.48	69.96	27.6
des	90.8	28.83	68.25	38.2
diffeq	62.6	26.66	57.41	-2.1
dsip	61.9	19.78	68.05	36.1
elliptic	107.1	42.76	60.02	6
ex1010	143.1	45.42	68.26	13.5
ex5p	168.2	41.43	75.37	55
frisc	129.6	42.82	66.96	-5.1
misex3	67.4	24.94	63.00	43.2
pdc	143.9	45.86	68.13	47
s298	130.81	45.81	64.98	50
s38417	75.46	30.69	59.33	29
s38584	118	40.51	65.67	38
seq	64.58	24.59	61.92	28
spla	109.54	38.29	65.04	50
tseng	131.1	45.51	65.07	16
ava	211.5	111.21	47.42	-0.1
Average	109.96	39.23	64.10	31.3

2D and 3D switch blocks distribution measured an average speed improvement of 31.3% and the 3D Mesh-based FPGA exploration methodology recorded counterproductive results for few benchmarks as illustrated in Table I. Nevertheless the 3D Tree-based FPGA measured consistent performance in all benchmark circuits.

VIII. 3D Tree-based FPGA Architecture Optimization

The experiments carried out in our laboratory and recent publications points out that the utilization of TSVs is actually very low in 3D FPGA with full vertical connectivity, which motivates us to develop an interconnect optimization module for Tree-based FPGA architectures [8]. By adopting Tree-based FPGA architecture and 3D technology, we strive for the feasibility of reducing the interconnect area requirement and power with only minor impact on speed by properly tailoring the structure and development strategy of partitioning and optimization of Tree-based multilevel programmable interconnect network using BFT topology as illustrated in Figure 11. To make 3D Tree-based FPGA more efficient in terms of design and manufacturing, it is essential to minimize the TSV count because TSV consumes more silicon area than horizontal interconnects and also reduces flexibility in placement and routing during physical synthesis [7]. The optimization experiments were performed individually for each netlist including architecture optimization section described in section IV. The architecture definition, partitioning, placement, routing and optimization is performed individually for each netlist listed in Table I. The TSV and architecture optimization are performed based on Rent's parameter p [10]. For the downward MSB (DMSB) interconnection network, the reduction in the number of inputs at level ℓ impacts level $\ell + 1$, since the number of DMSBs at level $\ell + 1$ is equal to the number of inputs at level ℓ. In the case upward interconnection network, reduction in

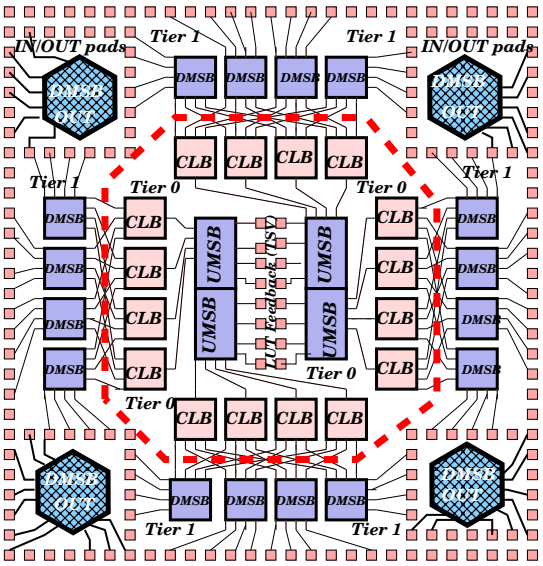

Fig. 11. 2 Tier 3D Tree-based FPGA cluster tile with 16 LUTS and BFT based interconnects

number of outputs at level ℓ impacts at level $\ell+1$ since the number of UMSBs at level $\ell+1$ is equal to the number of outputs at level ℓ. The optimization of upward and downward networks based on Rent's parameter [10] as follows.

$$N_{switch}(Tree) = N \times (k^p c_{in} + 2kc_{out}) \times \sum_{\ell=1}^{\log_k(N)} k^{(p-1)(\ell-1)} \quad (1)$$

The Tree level is represented as ℓ and k is the cluster arity, c is the number of in/out pins of an LB and IO is the number of in/out pins of a cluster located at level ℓ. The TSV count minimization methodology is implemented using Rent's parameter based iterative negotiation based 3D router program [8]. The aim is to find the best tradeoff between device routability and interconnect requirement of each application. The optimizer selects all 3D nets in the horizontal break point level of the Tree and optimize the number of vertical interconnects required between tier 0 and 1 of the horizontally partitioned 3D Tree-based FPGA. After completing the TSV count optimization at the break-point level, the interconnect optimizer choose the other levels above or below the break point levels on tier 0 and 1 to optimize the required number of routing resources in the upward and downward BFT based interconnection network in a random order.

Figure 11 presents the floorplan of the cluster with 2 levels including 16 LUTs. It also shows how we partition the Tree-based Butterfly-Fat-Tree interconnect network. The red dotted line indicate the horizontal partitioning of the Butterfly-Fat-Tree interconnect network. In this example, there are 84 nets in tier 0, that requires TSV based communication to tier 1. Out of 84 3D nets, 64 are inputs to LUTs provided by the downward interconnection networks (DMSBs) from tier 0 and other 20 are feedback from cluster LUTs as illustrated in Figure 11. The test chip we designed using 3D design process has 7 Tree levels with arity set to 4. It includes 16K LUTs and the horizontal break-point placed between levels 3 and 4 of the

TABLE II. INTERCONNECT OPTIMIZATION RESULTS

Tree Levels=7	Arity=4, Arch=4x4x4x4x4x4x4		
Tree-based Architecture Levels	3D Chip Active Layer	Optimized Rent 'p'	Optimized Area μm^2
Logic Blocks	Layer 1	–	93635273
Switch Level 0	Layer 1	0.67	2412
Switch Level 1	Layer 1	0.54	10800
Switch Level 2	Layer 1	0.66	37496
Switch Level 3	Layer 1	0.59	232128
$BreakPoint_{Horizontal}$	Horizontal Break Point (41%)		
Level 3 to 4	TSV Area=4423.68μm^2		
Switch Level 4	Layer 2	0.67	6072770
Switch Level 5	Layer 2	0.66	45553499
Switch Level 6	Layer 2	0.62	42139683
Average	-	0.63	-
Speed Degradation(Tree)	Horizontal=2.6% for 41% TSV reduction		
Speed Degradation(Mesh)	Horizontal=6.8% for 30% TSV reduction		

BFT based interconnect network. We have 4096 3D nets called cluster inputs pins and 1024 3D feedback networks pins. For a fully connected horizontally partitioned 3D Tree-based FPGA expected to have 5120 3D nets requires TSV communication excluding I/O pads in tier 0. The aim of TSV and interconnect optimizer is to find minimum number TSVs required between tier 0 and 1, and also optimize the interconnect area required to place and rout each MCNC application in other levels of the Tree. The optimizer will consider the same architecture with different Rent values p to find the minimum number of TSVs required to implement each application netlist. The purpose is to find for all benchmark circuits, the architecture with the smallest necessary TSV communications.

Table II presents the vertical interconnect optimization results. An average reduction of 40.1% TSVs and an average speed degradation of 2.6% recorded in these experiments. With help of vertical interconnect optimization, we reduced 2048 TSVs in our 2 tier 3D chip and at same time the impact of speed is also minimized. We conducted similar experiments for 3D Mesh-based FPGA using the methodology reported in [11]. Results presented in Table II. We observed a speed degradation of 6.8% for 30% reduction in TSV count. Unlike 3D Mesh-based FPGA, in Tree-based FPGA, we do not have to deal with 2D and 3D switch-blocks, but with only uniform switch types. Considerign the experimental results and methods used, we believe the 3D Tree-based FPGA is a consistent architecture to build more practical high performance 3D FPGA systems.

IX. TSV BASED FPGA THERMAL CONTROL

One of the main concerns in the design of 3D-ICs is heat desipation [13]. By stacking multiple active layers and increasing logic density, it become more difficult to remove the inter-tier heat. This problem is not so severe in our 3D stacked 2 tier Tree-based FPGA, since the heat produced in logic layer (tier 1) is easier to remove. However our aim is to build high density multi-stack 3D Tree-based FPGA. For this reason we developed an 3D thermal resistance mesh based multi-layer thermal model for FPGA based systems including TSV and power distribution network controlled heat transfer module. Using this module, the inter-layer temperature is optimized by considering spatial distribution of TSVs and PDNs. The 3D thermal model considers the impact of TSVs

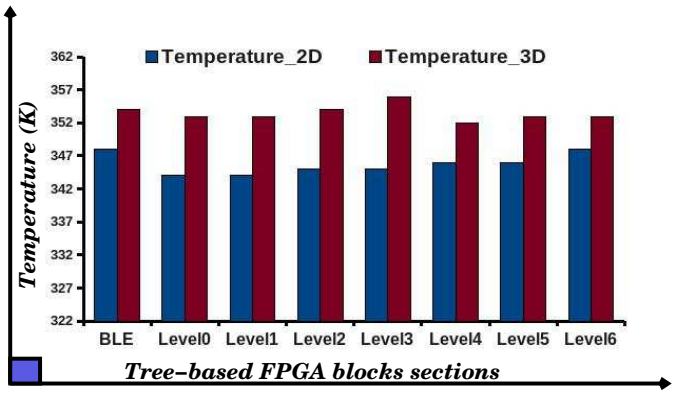

Fig. 12. Temperature values of functional units with and without Thermal TSV

material (Cu,Tungsten or doped Poly-silicon) while estimating the temperature profile. The effective thermal conductivity of active layers in 3D stacked chip is calculated by equation 2

$$k_{eff} = k_{cu} \times (TSV_{Area}) + K_{th} \times (Level_{BPArea} - TSV_{Area}) \quad (2)$$

In our 3D thermal experiments, we considered all FPGA blocks like BLE and interconnect levels range from 0 to 6 of the tree interconnect structure. We analyzed the thermal variation in 2D and 3D Tree-based FPGA using the 3D structure shown in Figure 5. The thermal analysis structure is configured using face-to-back configuration as illustrated in Figure 8. The 2D individual die level and inter-tier temperature profile extracted using 3D FPGA thermal model. The measured peak temperature of 2D heterogeneous Tree-based FPGA is 351K and average temperature is 346K. The temperature analysis of the 2 tier 3D Tree-based FPGA is presented in Figure 12. With our localized rearrangement of HBs and switch blocks along with spatial distribution of TSVs and PDNs, the peak and average temperature optimized at 357K and 351K respectively for 3D stacked horizontally partitioned Tree-based FPGA.

X. Conclusion

A systematic 3D physical design and implementation methodologies for Tree-based FPGA are presented. The issues associated with partitioning, TSV count and its impact on speed, design and manufacturing of 3D FPGAs are studied and presented. The study reveals the challenges in optimizing BFT based interconnect topology using horizontally partitioned 3D technology. We also present the management of TSV count and thermal control of 3D stacked Tree-based FPGA for guaranteed performance and yield. Therefor we believe that all the design and architecture styles presented in this paper can serve as a robust foundation for the design and manufacturing of even more practical 3D re-configurable systems based on horizontally partitioned Tree-based FPGA architectures.

References

[1] S M Trimberger, *Field Programmable Gate Array Technology*, Norwell, MA: Kulwer, 1994

[2] I Koun and J Rose,*Measuring the Gap between FPGAs and ASICs*, FPGA'06, Monterey, California, USA, February 2006.

[3] A DeHon, *"Reconfigurable Architectures for General-Purpose Computing"*, Ph.D dissertation, Department of Elect Engg and Computer Science, Massachusetts Institute of Technology, 1996.

[4] L Shang, A S Kaaviani and K Bathala, *Dynamic Power Consumption in Vertex-II FPGA Family*, in Proc ACM/SIGDA FPGA'10, pp. 157-164, 2002.

[5] E Ahmed and J Rose,*The Effect of LUT and Cluster Size on Deep-Submicron FPGA Performance and Density*, IEEE Transactions on VLSI Systems, pp:100-109, 2003.

[6] A Rahman, R Reif, *" System Level Performance Evaluation of Three Dimensional Integrated Circuits"*, IEEE Trans VLSI System, Vol 8, PP 671-678, Dec 2000.

[7] D H Kim, S Mukhopadhyay, and S K Lim, *"Through Silicon Via Aware Interconnect Prediction and Optimization for 3D Stacked ICs"*, Proc of ACM/IEEE Inter Workshop on System Level Interconnect Prediction, pp 85-92, 2009.

[8] V Pangracious, Z Marrakchi, E Amouri, and H Meherez, *Performance Analysis and Optimization of High Density Tree-based 3D Multilevel FPGA*, ARC 2013, pp 197-209, 2013.

[9] S Gupta, M Hilbert, S Hong, R Patti, *Techniques for Producing 3D ICs with High-Density Interconnect*, Tezzaron Semiconductor Naperville,IL 2005.

[10] Z Marrakchi, H Mrabet, U Farooq, H Mehrez:*FPGA Interconnect Topologies Exploration.* Int. J. Reconfig.Comp.2009

[11] K Siozios, A Bartzas, D Soudris.*Architecture Level Exploration of Alternative schmes Targeting 3D FPGAs: A Software Supported Methodology*, International Journal of Reconfigurable Computing,Vol2008.

[12] C Ababei, Y Feng, B Goplen.*Placement and routing in 3D integrated circuits*, IEEE Design & Test of Computers, vol.22, no.6, pp.520531,2005.

[13] V. Pangracious, Z. Marrakchi, E. Amori and H. Habib.*Physical Design Exploration of 3D Tree-based FPGA Architecture* IEEE/ACM GLSVLSI 2013, Paris France.

[14] S Lopez-Buedo, J Garrido, and E Boemo, *Dynamically inserting, operating, and eliminating thermal sensors of FPGA-based systems*, IEEE Transactions on Components Packaging Technology (CPM),vol. 25, no. 4, pp. 561566, Dec. 2002.

[15] S Velusamy et al., *Monitoring temperature in FPGA based SoCs*, International Conference on Computer Aided Design (ICCAD), San Jose, CA, 2005.

[16] J L Ayala, A Sridhar, V Pangracious, D Atienza, Y Leblebici *Through Silicon Via-Based Grid for Thermal Control in 3D Chips*, NanoNet pp.90-98 2009.

[17] D M Jang, C Ryu, K Y Lee, *Development and evaluation of 3-D SiP with vertically interconnected Through Silicon Vias (TSV)*, Proceedings of the 57th Electronic Components and Technology Conference (ECTC '07), pp.847-852, USA,May-June 2007.

[18] A Rahman, H Shi, Z Li, D Ibbotson and S Ramaswami, *Design and Manufacturing Enablement for Three-Dimensional (3D) Integrated Circuits (ICs)*, CICC pp. 1-8, 2012.

[19] L Madden, E Wu, N Kim, B Banijamali, K Abugharbieh, S Ramalingam and X Wu, *Advancing High Performance Heterogeneous Integration Through Die Stacking*, ESSCIRC, pp:18-24, 2012

[20] R Chaware, K Nagarajan and S Ramalingam, *Assembly and Reliability Challenges in 3D Integration of 28nm FPGA Die on a Large High Density 65nm Passive Interposer*, ECTC, pp; 279-283, 2012.

[21] M Lin, A EL Gamal, Yi-Chang Lu and S Wong, *Performance Benefits of Monolithically Stacked 3D FPGA*,Proceedings of the ACM/SIGDA 14th ISFPGA NY USA, pp 113-122, 2006.

[22] ITRS-2012, "International technology roadmap for semiconductors," [Online]. Available: http://public.itrs.net,March, 2012, pp:17-21

[23] V Pavlidis and E Friedman, *Interconnect-Based Design Methodologies for Three-Dimensional Integrated Circuits*,Proceedings of the IEEE, pp 123-140, Jan 2009.

[24] Andre DeHon, *Compact, Multilayer Layout for Butterfly Fat-Tree*, SPAA'00, pp:206-215, Bar Harbor, Maine, USA 2000.

[25] Cha-I Chen Bau-Cheng LEE and Juinn-Dar Huang, *Architectural Exploration of 3D FPGAs Towards A Better Balance Between Area and Delay*, DATE11, 2011.