

# **Impact of 3D IC on NoC Topologies: A Wire Delay Consideration**

Mohamad Jabbar, Dominique Houzet, Omar Hammami

# **To cite this version:**

Mohamad Jabbar, Dominique Houzet, Omar Hammami. Impact of 3D IC on NoC Topologies: A Wire Delay Consideration. DSD 2013 - Digital System Design: Architectures, Methods and Tools, Sep 2013, Santander, Spain. pp.68-72. hal-00938984

# **HAL Id: hal-00938984 <https://hal.science/hal-00938984>**

Submitted on 29 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mohamad Hairol Jabbar Department of Computer Engineering, Unitversiti Tun Hussein Onn Malaysia, 86400, Batu Pahat, Johor,Malaysia hairol@uthm.edu.my

Dominique Houzet GIPSA-Lab, Grenoble INP, 11 Rue des Mathematiques 38402, Grenoble,France dominique.houzet@grenoble-inp.fr

Omar Hammami ENSTA Paristech, 828, Boulevard des Maréchaux, 91762, Palaiseau, France omar.hammami@ensta-paristech.fr

*Abstract*—In this paper, we perform an exploration of 3D NoC architectures through physical design implementation based on two tiers Tezzaron 3D technology. The 3D NoC partitioning is done by dividing the NoC's datapath component into two blocks placed in the two tiers. Two Stacked NoC architectures namely Stacked 3D-Mesh NoC and Stacked 2D-Hexagonal NoC developed based on this partitioning strategy are analyzed by comparing their performances with Stacked 2D-Mesh NoC and classical 2D-Mesh and 3D-Mesh NoC. In order to measure the impact of wire delay on performance, two technology libraries (130 nm and 45 nm) representing old and advanced technologies have been used for the performance analysis. Results from physical implementations show that in advanced technologies such as 45 nm and below, the performance of Stacked 2D NoC topologies with datapath partitioning method have better performances compared with traditional 2D/3D Mesh topologies and Stacked 3D Mesh topology. We advocate here that with stacking there is no need for 3D NoC topologies for advanced 2-tier 3D IC and this is also confirmed for multistage networks like butterfly.

Keywords—3D NoC Architecture, Exploration, Network on Chip, Partitioning, Physical design

#### I. INTRODUCTION

As moving to sub-20 nm CMOS technology poses great design and manufacturing challenges, 3D integration [1] is increasingly seen as a solution to those challenges for designing complex system on chip [2]. Global interconnect wire is one of the primary concern for advanced process technology (65 nm and below) (Figure 1) that has substantial contribution to the wire delay as well as power consumption even with the repeaters. By stacking dies or wafers, the performance can be increased due to reduction in interconnect length and so is power consumption due to the reduction in number of repeaters along the wires.

Performance improvement of 3D integration can be more prominent when compared with shrinking transistor technology [3]. Implementing the 2D design in 3D architecture on the same process technology could provides higher performance benefit than CMOS migration to the next process technology. Stacking multiple dies will also reduce the total footprint of a chip making it very suitable for mobile devices. However, several challenges such as thermal, yield, cost, design tools and testing of 3D architecture must be overcome before 3D IC technology can be widely adopted as a mainstream technology [4].

This paper presents an exploration of 3D NoC architectures through physical design case studies. Our motivation is that we want to explore different partitioning strategies from the previous reported works for 3D NoC architecture and then evaluate their performance accurately based on layout-level routed netlist. The contributions of this work are as follows:





- Propose a partitioning strategy to optimize NoC architecture performance based on stacking router's datapath components into two blocks on two tiers.
- Study the wire delay impact on the 3D NoC architectures by performing physical implementation using two standard cell libraries, 130 nm and 45 nm.

## II. RELATED WORKS

Many issues in 2D NoC architecture and design have been studied over the past years covering various aspects such as design flow, implementation evaluation and design space exploration. However, research in 3D NoC is still new and many issues remain unexplored especially in real design and implementation. Design space exploration of 3D NoC topologies through cycle accurate simulation have been performed showing the benefits of 3D design in terms of throughput, latency and energy dissipation for mesh-based and tree-based NoC architecture [5]. In [6], zero load latency and power consumption analytical models of various 3D NoC topologies have been evaluated proving the advantages of combining 3D IC with 3D NoC architecture. We base upon this literature to investigate further the results by doing analysis from physical design implementation results. Another work [7] proposed a novel 3D router architecture by decomposing the router into different dimensions to provide better performance over other 3D NoC architectures. We differ from the previous reported works as we focus on partitioning 3D NoC architectures and evaluate their performance through layout-level netlist for more accurate analysis of wirelength, timing, area and power consumption.

Several experiments have been conducted investigating the performance of 3D architectures based on the results from physical design implementations. Work in [8] has studied different partitioning styles for implementing 3D multicore architectures namely core level, block level and gate level showing that TSV capacitance, EDA tools and timing optimization methods have strong impact on the performance of the final 3D architecture. In [9], they showed that 3D architecture could lose or reduce its benefit due to the tools

inability to perform 3D-aware optimization. On the other hand, larger circuits tend to gain more improvement from 3D architecture over its 2D counterpart for advanced technology such as 45 nm node. In [10], the study of different 3D placement methods on the performance of three 3D architectures showed that true-3D placement method produces the highest performance improvement over other methods at old technology (130 nm) indicating the importance of 3D-aware tools to obtain maximum benefits of 3D integration. However, no previous work has been presented with detailed performance evaluation on various physical design metrics (wirelength, timing, impact of wire length) of 3D NoC architecture in particular with 3D Meshbased NoC architecture.

#### III. 3D IC TECHNOLOGY

The 3D integration technology we used is based on Tezzaron [11] that uses TSV for peripheral IOs and microbumps for inter die connections. The two-tier 3D stacking method is based on wafer-to-wafer bonding, face-to-face method with via-first approach as illustrated in Figure 2. The inter-die microbumps provide high interconnection density up to 40,000 per mm<sup>2</sup> without interfering to FEOL (front-end-ofline) device or routing layers. It is also possible to implement four tiers by stacking through back-to-back using TSV of the two face-to-face stacking in order to have higher design complexity but it will not be covered in this paper.



Figure 2: Cross section of Tezzaron 3D IC technology with corresponding parameters

In order to analyze performance of 3D NoC architectures in advanced technology, we have chosen 45 nm standard library from ST Microelectronic [12]. We use similar 3D structure for inter-tier connections using microbumps as in Tezzaron technology but we replace the 130 nm technology of Global Foundries with 45 nm ST Microelectronic standard library. The 45 nm technology used in this study has seven metal layers where metal seven is used for bonding and the routing is limited until metal six.

#### *A. Design Flow Based on 2D EDA Tools*

The 3D design flow is developed based on the 2D EDA tools. This flow is made possible with the Tezzaron 3D technology using mirobumps for inter-tier connections. These microbumps have negligible delay for the inter-tier connection and thus we can perform 3D timing analysis at post-synthesis stage without any inaccurate delay estimation of inter-tier connection. Post-synthesis static timing analysis (STA) for each tier is done separately before 3D timing analysis is performed. In order to perform 3D timing analysis

at post-synthesis stage, we create a top level netlist that instantiate both tiers and connect them using inter-tier wires that represent microbumps. Using the generated timing constraints, timing optimization is carried out using 2D place and route tool for each tier separately. For post-route 3D timing analysis, we create the top level netlist as in the postsynthesis step and feed the SPEF file of each tier into Synopsys PrimeTime for timing and power analysis. The parasitics for the microbumps are ignored due to their negligible delay.

### IV. BASELINE NOC ARCHITECTURE

#### *A. Router and Network Interface Architecture*

The router and network interface architectures are standard architectures used for mesh NoC. In this experiment, we did not include the processor core in order to make the experiment easier. Also, if we include the processor cores, the results will be more significant (the benefit of the partition will be higher) because of the increased inter-router links.

#### *B. Baseline 3D Mesh NoC*

In this architecture, the 3D NoC is implemented on two tiers where each tier has identical blocks as shown in Figure 3. This is the straightforward extension of 2D Mesh NoC architecture to 3D Mesh NoC where we just take a copy of a tile (a router and a network interface) and put it on top of each tile. Compared with the area of 2D Stacked Mesh NoC, this architecture has slightly more area due to the additional ports for vertical connections. This 4x2x2 Mesh NoC architecture is based on a 3D router architecture that has vertical links for inter-tier connections between routers. These physical vertical links shown in red color are based on the logical vertical links in each 3D router.

#### V. EXPLORATION OF 3D NOC ARCHITECTURES *A. 3D NoC Partitioning*

In this section, we describe the partitioning method to be used for the next 3D stacked NoC architectures. The FIFO buffer is dominating silicon area in the NoC architecture. Thus, it is a good approach to partition it into two tiers. Other datapath components are also partitioned into two tiers at bitlevel. For example, for the 32 bit FIFO size, the resulting implementation will be 16 bits per tier. For the non-datapath components such as routing logic, arbitration logic and FIFO control, we place them on each tier by trying to balance the area of both tiers. Figure 4 illustrates this partitioning method with respects to 2D and baseline 3D Mesh NoC architectures. Rather than using automatic tools such as HMetis to partition the design, we focus on dividing the datapath manually into two parts and place them into two tiers in order to preserve the homogeneous properties of tile block architecture. Another reason for not using this automatic tool is because the tool also tries to optimize the nets between gates in the netlist with no capability of 3D placement meaning that logic cells can be interchangeably partitioned into the two tiers which will eventually affect the 3D timing path.



Figure 3: 3D Mesh NoC design a) block diagram b) floorplan c) routed layout

#### *B. 3DNoC1: 3D Stacked Mesh NoC*

The first architecture designed with the partitioning method is depicted in Figure 4. Here, rather than stacking the tiles on top of each other, we map the 3D NoC on the 2D layout and then partition it into two tiers. As shown in Figure 6(a), the green links represent logical vertical connections between 3D routers while the physical vertical links in orange color are basically the 2D logical links. By doing this, the area is slightly increased compared with the 3D Mesh NoC but reduced compared with the 2D one. However, this partitioning method requires higher number of inter-tier connections than pure 3D Mesh NoC. One disadvantage of this structure is that the inter-router wire links are not equal between all routers because vertical wire links are longer than other links.



Figure 4: Partitioning method for the 3D stacked NoC architecture

#### *C. 3DNoC2: 3D Stacked Hexagonal NoC*

Due to unequal inter-router wire link lengths in the 3D Stacked Mesh NoC architecture because of the logical vertical links (green lines in Figure 6(a)), we proposed a new topology having same length of inter-router physical links called hexagonal topology shown in Figure 8 (a). We used the here the same datapath partitioning method (cf. section 5.1). Previous work [13] had proved that hexagonal topology is the most efficient topology and theoretical exploration of addressing, routing and broadcasting in hexagonal mesh architecture has also been explored.

#### *1) Packet Routing*

Routing is illustrated in Figure 5. Basically the packets will be first routed through X direction and then to Y direction to reach the destination. However, in the case of a router with a diagonal link in the same direction as Y direction, the packets will be routed through this diagonal link instead of X axis link. Therefore, from the diagram, the packets will be routed from router 00 to 33 through router 11, 12 and 23. The diameter of the Hexagonal NoC can be formulated as  $d = (x-1) + (y-1) - (x/2)$ , where *x* is the number of hops in X axis and *y* is the number of hops in Y axis. This hexagonal routing is a dimension ordered routing which is thus deadlock free.



Figure 5: Hexagonal routing block diagram



Figure 6: 3D Stacked Mesh design a) block diagram b) floorplan c) routed layout

#### *2) Physical Implementation*

The physical size of the tiles is determined by measuring the distance between tiles such that the distance between the six neighboring tiles is equal. This is to make sure that this square floorplan area is identical to the original hexagonal shape. Although it is possible to create hexagonal floorplan in SoC Encounter, we choose to adopt rectangular floorplan. As shown in Figure 7, the rectangular floorplan for the hexagonal architecture can be carefully arranged such that the interrouter links are equal and use only vertical and horizontal links, and thus we avoid the use of diagonal wires in the case of an hexagonal shape having four diagonal edges. We adopted the equation of  $(a/2)^2 + b^2 = c^2$ , where *a* is the tile's height, *b* is tile's width and *c* is the physical direct distance between the two tiles to determine the size of each tile. We first fix the value of *a* and then find the value of *b* such that *c* is equal to *a* at the same time meeting initial target utilization. We also have derived mathematical formulation proving that the surface area of the square floorplan is identical to the original hexagonal structure. Let's say *a* equal to 579 µm, following the equation above will obtain the value of *b* equal to 500 µm and *c* equal to 577 µm with the initial target utilization of 60%. To compare the diameter for both topologies, consider an example of a 4 x 5 network. The diameter for 3D Mesh NoC is 6 while for 3D Hexagonal NoC is 6 but Hexagonal NoC has shorter inter-router wire links benefit.

#### VI. EXPERIMENTAL RESULTS

For older technology such as 130 nm and above, wire length effect is not significant and the delay in the critical paths is mostly determined by the delay of the gates. As shown in Table I and Figure 9, the 3D NoC architectures do not benefit in terms of speed and power consumption. The power consumption is even higher in 3D architectures due to the additional gates as well as the increased wirelength. In this study, we used simple partitioning method to partition the 2D design into 2 tiers. However, some studies have shown that automatic partitioning tools could provide performance improvement over 2D architecture even using old technology such as 130 nm and 180 nm [14]. Partitioning is very important in 3D design primarily for old technology. Using automatic partitioning tool such as hMetis [15] helps to improve the performance of 3D architecture although it is still not significant because the tool try to optimize the connections between gates in the synthesized netlist but is not able to perform in-place 3D optimization during place and route as in usual 2D optimization. At 45 nm, automatic partitioning tools can provide higher performance improvement for the 3D architecture than for old technology.

#### *A. Analysis on the Impact of Wire Delay*

As for wire delay, older technology nodes (such as 130 nm) do not have significant wire delay on the performance. The critical path for all designs in this study is located within the tile block (from bottom tier to top tier in 3D stacked architecture) except for 3D Mesh NoC architecture where its critical path is between two routers. Looking at the 3D critical paths for all 3D NoC architectures in this study, the ratio of wire delay is about 3% of the total critical path delay. For comparison, the wire delay in 2D architecture using the same process technology is about 5.7% of the total critical path delay and thus we can generally conclude that 3D architecture in this technology will not offer any benefit in terms of speed.

However, there is still an opportunity to gain benefit from 3D architecture by optimizing partitioning method as demonstrated by several works previously using older technology nodes [10] although the results is not very significant compared with ideal improvement we should get. Additionally, analyzing the critical path delay for the 2D architecture using 45 nm technology indicates that wire delay is about 1% due to very small area. We expect to see larger portion of wire delay in the critical path for 2D architecture with larger design.



Figure 7: Rectangular floorplan for hexagonal architecture showing the inter-router links



Figure 8: 3D Stacked Hexagonal NoC a) block diagram b) floorplan c) routed layout

Table I: Physical implementation results for 3D NoC architectures in 130 nm technology

Parameters	3D Mesh NoC	3D Stacked Mesh NoC	3D Stacked Hexagonal NoC
Vertical connections	1763	6261	7255
Core area $(mm2)$	3.24	4.37	5.43
Total wirelength (m)	12.48	14.01	17.03
Longest path (ns)	4.20	4.60	4.73
Power $@$ 333 MHz (W)	1.44	1.25	1.40



Figure 9: Performance comparison of 3D NoC architectures over 2D NoC in 130 nm technology

Table II: Physical implementation results of 3D NoC architectures in 45 nm technology

Parameters	3D Mesh NoC	3D Stacked Mesh NoC	3D Stacked Hexagonal NoC
Vertical connections	1763	6261	7255
Core area $(mm2)$	0.79	0.91	1.01
Total wirelength (m)	5.5	5.9	6.5
Longest path (ns)	3.23	3.33	3.59
Power @ 333 MHz (W)	0.23	0.24	0.26

*B. Extrapolation for Advanced Technology* 

For designs using 45 nm used in this study, the 3D architectures still do not provide any improvement over its 2D design as shown in Table II and Figure 10. However, it shows a reduction trend of the gap between 3D and 2D architectures compared with the results in Figure 9 using 130 nm technology. If we look at the area, we can see that this design consumes very small area (less than  $1 \text{ mm}^2$ ) and this is the primary reason why there is no improvement obtained using 45 nm technology. Previous work have demonstrated for large designs (about  $36 \text{ mm}^2$  in 2D architecture), substantial performance improvement (75% reduction in longest path delay) that could be achieved over 2D architecture using the same 45 nm technology because wirelength becomes significant [9]. Table III shows the extrapolation of wire delay for 22 nm technology based on the critical path wire delay in 45 nm and on the data from ITRS 2007 interconnect report for global wire without drivers (1.02ns (45nm), 3.3ns (22nm), 5.9ns (16nm)). This extrapolation is intended to show that when the design used is realistically large, we will see improvement for the proposed hexagonal NoC topology in stacked 3D architecture compared to the other solutions. The gate delay value for 22 nm is assumed to improve two times over 45 nm technology because it is two technology generations from 45 nm and the tile area (and thus the interrouter wire length) is assumes to be 3 mm x 3 mm for the 3D Mesh NoC considering the area of commercial grade LEON3 processor [9]. From the 3 mm inter-router wirelength of 3D Mesh NoC, we calculate the wire length for 2D Stacked Mesh and 3D Mesh as follows:

From the rectangular area equation for the hexagonal floorplan,

$$
\left(\frac{a}{2}\right)^2 + b^2 = c^2
$$

And  $c = a$  (because equal inter-router links), thus,

$$
b = \sqrt{a^2 - \left(\frac{a}{2}\right)^2} = \sqrt{3} * \frac{a}{2} = 0.866a
$$

For 2D Stacked Mesh,

$$
x2 = a * b
$$

$$
x = \sqrt{0.866}a = 0.93a
$$

For 3D Mesh (double of 2D Stacked Mesh),  $y = 2 * 0.93a = 1.86a$ 

where *a* is the inter-router length for 3D Stacked Hexagonal NoC, *x* and *y* is the new inter-router length for 2D Stacked Mesh and 3D Mesh respectively. The wire length for 3D Mesh and 3D Stacked Mesh is equal because 3D Stacked Mesh has half the area of 3D Mesh but has double the interrouter length for 3D logical vertical links due to the 2D mapping of the 3D Stacked Mesh. This extrapolation is simplified by ignoring the router area impact due to the different number of ports for different topologies which is 4, 5 and 6 ports for 2D Mesh, 3D Mesh (also 3D Stacked Mesh) and 2D Hexagonal respectively. As can be seen from the table, the wire delay is becoming more significant for 16 nm technology and thus it will have strong impact on the critical path delay especially for 3D Mesh NoC and 3D Stacked Mesh NoC (because of logical vertical links between routers) since it has longer inter-router wire links. The 2D Stacked Mesh NoC outperforms the 3D Mesh NoC and 3D Stacked Mesh

NoC in the wire delay and eventually the total delay. However, the 3D Stacked Hexagonal NoC is shown to have better improvement than 2D Stacked Mesh in terms of network latency because it has a lower diameter compared with 2D Stacked Mesh and thus it will benefit for applications running on large networks.

### *C. Extending for Other NoC Topology*

The results from this experiment can also be applied to other topologies such as multi-stage interconnection and butterfly network. Previous work has proposed to partition the butterfly network by folding it into several tiers [16]. As has been explained in previous sections for the hexagonal topology, the same method can also be used for butterfly topology, meaning that partitioning it into two tiers (which can be referred to 3D Stacked Butterfly) could improve its performance due to the shorten wire links between stages.



Figure 10: Performance comparison of 3D NoC architectures over 2D NoC in 45 nm technology





#### VII. CONCLUSIONS

This paper has presented an exploration of 3D NoC architectures through physical design case studies. The performance of these 3D NoC architectures, namely 3D Stacked Mesh NoC and 3D Stacked Hexagonal NoC has been analyzed by comparing with the 2D Mesh NoC and also traditional 3D Mesh NoC architecture. For advanced technologies such as 45 nm and beyond, the 3D NoC architectures based on this partitioning method show better performance than traditional 3D Mesh NoC architecture due to significance of wire delay effect. Future work will extend this study to manycore MPSoC and 3D EDA tools implementation.

#### **REFERENCES**

- [1] J. U. Knickerbocker, P. S. Andry, B. Dang, R. R. Horton, M. J. Interrante, C. S. Patel, R. J. Polastre, K. Sakuma, R. Sirdeshmukh, E. J. Sprogis, S. M. Sri-Jayantha, A. M. Stephens, A. W. Topol, C. K. Tsang, B. C. Webb, and S. L. Wright, "Three-dimensional silicon integration," *IBM Journal of Research and Development*, vol. 52, no. 6, pp. 553–569, 2008.
- [2] "ITRS Report Interconnect," 2011. [Online]. Available:
- http://www.itrs.net/home.html.
- [3] K. Nomura, K. Abe, S. Fujita, Y. Kurosawa, and A. Kageshima, "Performance Analysis of 3D-IC for Multi-Core Processors in sub-65nm CMOS technologies," in *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, 2010, pp. 2876–2879.
- [4] J. Li and C.-W. Wu, "Is 3D integration an opportunity or just a hype?," in *Design Automation Conference (ASP-DAC), 2010 15th Asia and South Pacific*, 2010, vol. 978, pp. 541–543.
- [5] B. S. Feero and P. P. Pande, "Networks-on-Chip in a Three-Dimensional Environment: A Performance Evaluation," *Computers, IEEE Transactions on*, vol. 58, no. 1, pp. 32–45, 2009.
- [6] V. Pavlidis and E. Friedman, "3-D Topologies for Networks-on-Chip," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 15, no. 10, pp. 1081–1090, Oct. 2007.
- [7] J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, V. Narayanan, M. S. Yousif, and C. R. Das, "A novel dimensionally-decomposed router for on-chip communication in 3D architectures," in *Proceedings of the 34th annual international symposium on Computer architecture*, 2007, pp. 138–149.
- [8] Y.-J. Lee and S. K. Lim, "Timing analysis and optimization for 3D stacked multi-core microprocessors," in *3D Systems Integration Conference (3DIC), 2010 IEEE International*, 2010, pp. 1–7.
- [9] M. Pathak, Y.-J. Lee, T. Moon, and S. K. Lim, "Through-siliconvia management during 3D physical design: When to add and how many?," in *Computer-Aided Design (ICCAD), 2010 IEEE/ACM International Conference on*, 2010, pp. 387–394.
- [10] T. Thorolfsson, G. Luo, J. Cong, and P. D. Franzon, "Logic-onlogic 3D integration and placement," in *3D Systems Integration Conference (3DIC), 2010 IEEE International*, 2010, pp. 1–4.
- [11] R. S. Patti, "Homogeneous 3D Integration," in *Three Dimensional System Integration: IC Stacking Process and Design*, A. Papanikolaou, D. Soudris, and R. Radojcic, Eds. Springer US, 2011, pp. 51–71.
- [12] S. Microelectronic, "45 nm ST Microelectronic standard library." [Online]. Available:
- http://cmp.imag.fr/products/ic/?p=STCMOS040. [13] A. D. Malony, "Regular processor arrays," in *Frontiers of Massively Parallel Computation, 1988. Proceedings., 2nd Symposium on the Frontiers of*, 1988, pp. 499–502.
- [14] T. Thorolfsson, N. Moezzi-Madani, and P. D. Franzon, "Reconfigurable five-layer three-dimensional integrated memoryon-logic synthetic aperture radar processor," *Computers & Digital Techniques, IET*, vol. 5, no. 3, pp. 198–204, 2011.
- [15] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar, "Multilevel hypergraph partitioning: applications in VLSI domain," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 7, no. 1, pp. 69–79, 1999.
- [16] O. Hammami, A. M''zah, and K. Hamwi, "Design of 3D-IC for butterfly NOC based 64 PE-multicore: Analysis and design space exploration," in *3D Systems Integration Conference (3DIC), 2011 IEEE International*, 2012, pp. 1–4.