A Review of the Design Challenges for the 3-D on Chip Network Paradigms

Neha Jain
M.Tech Scholar, CSE
GITS, Udaipur, INDIA

Mayank Patel
Department of CSE
GITS, Udaipur, INDIA

ABSTRACT
As feature sizes continue to shrink and integration densities continue to increase, interconnect delays have become a critical bottleneck in 2D NoC performance. The upcoming decades will require a change from mere transistor scaling to novel packaging architectures such as the vertical integration of chips referred as 3D integration. 3D silicon integration technologies have provided new opportunities for NoC architecture design in SoCs enabling the design of complex and highly interconnected systems in reduced space providing higher efficiency compared to 2D integration. The next challenge in front of researchers in the domain of NoC is to use NoC architecture as the backbone of the upcoming generation of 3D chips. Multiple design issues have to be addressed in this respect such as high chip temperature due to increasing power density leading to large interconnect-delays, lack of design methodologies, large area covered by vertical interconnects, problems related to optimally determining tier assignments and the placement of switches in 3D circuits. In this paper, we tried to exhibit and summarize the prevalent generic 3D NoC design issues highlighted by various recent research publications in the domain of NoC.

Keywords
3D ICs, Through Silicon Vias, Thermal, CMP

1. INTRODUCTION
As per the Moore's Law, the packing density in deep sub-micron technology doubles every 18 months, leading to increased research interests in the domain of network-on-chip (NoC) interconnection paradigm that provides flexibility, scalability and suitability in designing efficient SoCs. According to the International Technology Roadmap for Semiconductors (ITRS) [44] before the cessation of this decade we will be witnessing the era of a billion transistors on a singlechip. A hint of such a development means thatin the near future we probably have devices with such complexfunctions ranging from mere mobile phones to adaptable mobile-devices capable of controlling satellite functions. But developing suchkind of chips is not a trivial task as the number of on-chip transistors increases significantly and so does the complexity of integrating them. Today's SoCs use shared or dedicated buses to interconnect the communication of on-chip resources [1]. However, these buses are not scalable beyond a certain limit leading to the design productivity gap. Many other factors like heterogeneous components, power management and embedded software, wire delays, signal integrity, central arbitration and more design choices are making the entire design process more time consuming and complex.

In this scenario, the interconnect infrastructure will become a bottleneck for the development of billion transistor chips [1]. In order to alleviate complex communication problems and to replace global interconnects 2D NoC designs have proved to be beneficiary up to an extent. The design of 2D NoCs has been examined from various research publications of the aspects, such as performance, power and reliability [23-28].

The onset of three-dimensional (3D) stacked technologies provides a new horizon for on-chip interconnect design. Three-dimensional integrated circuits (3D ICs), contain multiple layers of active devices stacked onto each other, have the potential for enhancing system power/performance characteristics [10-14]. 3D ICs allow for performance enhancements even in the absence of scaling [11][15].

The amalgamation of two emerging paradigms, NoC and 3D IC, allows for the creation of new structures that enable significant performance enhancements over more traditional solutions [2].

[45] has proposed a heuristic approach and shown that 3D NoC performs better in terms of consumption of dynamic communication energy in comparison to 2D NoC with the same number of nodes.

[18] has proposed symmetric noc architecture where the nodes are grouped into multiple layers and are stacked on top of each other. Despite of simplicity, this architecture has a major inherent drawback as it does not exploit the beneficial attribute of a negligible inter-wafer distance in 3D chips.

[19] Introduces a new architecture called ciliated 3D Mesh. In a ciliated 3D Mesh network, each switch contains at most 5+k ports (one for each cardinal direction, two for up and down (one either up or down in two layer 3D mesh) and one to each of the k IP blocks.

Butterfly fat tree (BFT) [20], [21] and the generic fat tree, or SPIN [22] is the two types of tree-based interconnection networks that have been considered for NoC applications. According to [19], considerable enhancements can be achieved when these networks are instantiated in a 3D IC environment. Unlike the work with mesh-based NoCs, any new topologies for tree-based systems were not proposed instead the [22] presents an achievable performance benefits by instantiating already existing tree-based NoC topologies in a 3D environment.

One of the most important concerns in designing 3D NoC chips is the choice of suitable routing algorithm.

Most of the 3D NoC systems are based upon the Dimension Order Routing (DOR) XYZ algorithm which routes flits first along the X dimension, then along the Y and finally the flit is routed along the Z dimension to reach its destination. In [11] LA-XYZ (look ahead- XYZ) routing algorithm has been proposed.[12] Extends the routing algorithm BDOR designed for 2D-NoC.[13] Presents a novel fully adaptive and fault-tolerant routing algorithm for Network-on-Chips (NoCs) called Force-Directed Wormhole Routing (FDWR).

3D NoCs promise significant benefits but also impose new constraints and limitations.

The rest of the paper is organized as follows. Section II discusses motivations behind the 3D NoC; Section III covers the details of the existing design challenges of 3D NoC. Lastly, in Section IV we conclude.

2. MOTIVATIONS BEHIND 3D NOCs

The traditional 2D on-chip system, offers lower performance as the number of on-chip cores are increasing at an exponential rate along with restricted floor-planning choices. Besides, 2D ICs have larger die size in multiprocessor implementations with difficulty in clock distribution.

The 3D IC is a cutting edge technology in which multiple 2D chip layers are stacked vertically via layer-to-layer interconnections. Containing multiple layers of active devices, it can be predicted that 3D NoCs have the potential of enhancing system performance.

Industry and academia all over the world are currently working on the development of 3D system integration technologies of which some of the prominent ones are chipstacking, transistor stacking, die-on-wafer stacking, andwafer-level stacking[33]. Industrial users are therefore working in the areas of high-density memories, high performance processors and real time image processing and aerospace applications.

With the shrinking of the VLSI technologies (3D ICs) [34] have proved as an attractive option for overcoming the barriers in interconnect scaling, thereby offering an opportunity to continue performance improvements using CMOS technology.

3D integration technologies offer many benefits for future microprocessor designs. Such benefits include:

(1) The reduction in interconnect wire length, which results in improved performance and reduced power consumption;
(2) Diameter reduction, which results in the less number of hops traversed by the flits that leads to an improved throughput and reduced latency; [4] ONLY
(3) Improved memory bandwidth, by stacking memory on microprocessor cores with TSV connections between the memory layer and the core layer;
(4) The support for realization of heterogeneous integration, which could result in novel architecture designs highly apt for the application-specific NoCs;
(5) Smaller form factor, which results in higher packing density and smaller footprint due to the addition of a third dimension to the conventional two dimensional layouts and potentially results in a lower cost design [3].

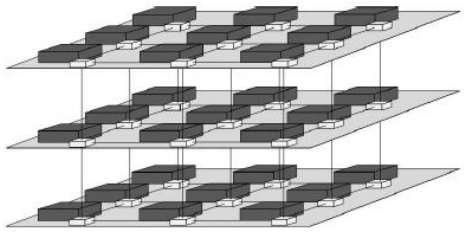

Fig.1: 3D IC model with NoC communication structure

Along with all these positive aspects there exist several challenges in integrating the 3D technology with the NoC framework which need to be addressed.

3. DESIGN CHALLENGES OF 3D NoCs

While the potential advantages of the 3DIC are certainly compelling, the following challenges have curtailed its adoption as a mainstream technology:

3.1 Thermal Management and Effective Cooling Challenges

Thermal management is one of the important issues of 3D IC integration. For it effective thermal management methodologies and solutions are needed to make widespread use of 3D IC integration. The thermal analysis is important with respect to thermal and stress distribution; efficient layout modification which enhances the circuit reliability. Thermal management in 3D stacks is critical for maintaining required reliability, performance, and power dissipation targets [5].

Subsequently, cooling of 3D assemblies is also difficult as the power per unit area increases and the heat produced must be conducted through multiple chips, often with poor thermal interfaces. Although it is possible to introduce coolants within a thick 3D structure to handle very high power levels [35], such an approach is complex and requires thick cooling structures within the stack to bring in a sufficient fluid volume. Thus every effort should be made to remove heat from the back of a chip stack as is currently done for single high-power chips.

Therefore, new cooling techniques must be innovated to ensure that the chip can operate at high temperatures and can achieve performances and reliability.

3.2 Design Methodologies and Circuit Architecture Challenges

The electrical integrity of devices and circuits must be preserved during the 3DIC fabrication process. One critical issue is thermal cycling during 3DIC fabrication, which can degrade device performance. New processes must be established to precisely align and interconnect the multiple device layers. To obtain the optimal circuit benefits the alignment has to be in the order of that in the critical layers; if that cannot he achieved, the resulting trade-offs must be investigated. 3D design methodologies are still needed to reduce coupling between TSVs (body contacts reduces the current loop so it reduces the mutual inductances) and for clock and power distribution networks for 3D ICs. Also, more investigations are required for analyzing the effect of TSVs on signal integrity, power integrity and delay. The 3-D integration also provides opportunities for new circuit architectures and to maintain desirable performance levels.

3.3 Minimizing Complexity

3D circuits are inherently complex, from- the standpoint of both design and fabrication. New design tools will be required to optimize interlayer connections for maximized circuit performance and the process fabrication complexity must be minimized for manufacturing ability and yield. A set of 3Dspecific tools for standard-cell placement, global routing, and layout was reported in [29][30].A reliability tool specifically targeted for 3Dtechnology is also available that helps in achieving “increased locality” as many more neighbors could be communicated within a few minutes of reach[30].

3.4 Managing Power

The increased circuit density inevitably leads to increased power density and new schemes may be required to enable power dissipation from multiple device layers in order to minimize thermal gradients and local heating [31]. As silicon process technology, driven by improved performance requirements, migrates to low resistivity metals, the thermal properties of these materials play a significant role in determining the temperature build-up, power and performance of the chip [32]. The temperature rise underneath the devices is much larger than the near interconnects because the heat dissipation across the buried dielectric layer is poor. This temperature rise has a direct impact on the performance of the system. For 3D ICs one solution is to place the large, highly loaded cores near the lower-most plane next to silicon substrate.

3.5 Electrical Modeling Challenging

The purpose of modeling interconnections is to extract equivalent circuits that describe the electrical characteristics of given interconnection structures. A major difficulty in modeling 3-D interconnections comes from the need to obtain the entire coupling model of a large number of 3D interconnections. Furthermore, for accurate electrical design 3-D NoC interconnection model should cover a sufficiently wide frequency range [7].

Unlike conventional interconnects, a TSV, is a more general structure, surrounded by the silicon substrate that has a finite resistivity. This finite resistivity results in a nonlinear capacitance and resistance parasitics that are not present in conventional interconnect and a large capacitance that is highly dependent on TSV relative position to nearby body contact. Therefore, in case of modeling TSV interconnections, the silicon substrate should be considered as another source of loss. Clearly, TSV modeling is extremely challenging when a large number of TSV interconnections are involved. [36]. One of the problems being researched in this domain is the modeling of multi-TSV arrangement in a 3-D system and their time-domain co-simulation [7].

Most of previous work characterizes specific TSV structures but they do not address general multi-TSV problems [7]. Other related work proposes a model but no closed-form expressions are given for the TSV parasitic calculations [37]. In other words, there has been no related work in TSV macro-modeling. Also, few works addresses the multi-stacked TSV modeling problem [7] these issues should be carefully evaluated.

3.6 CAD/EDA Tools Availability

New tools that consider thermally aware physical design implementations, most importantly at the architecture and SoC level are crucial to the success of 3D as thermal issues are exacerbated in 3D implementations [38]. To justify the cost and complexity overhead of 3D technology, it is essential to study the benefit of 3D early in the design cycle. This requires strong linkage between architecture level analysis tools and 3D physical planning tools. Most of the advantages of 3D will be utilized with new system architectures and physical implementations [5][7].

Therefore, the tools to aid 3D implementation must also operate at the higher level. Therefore, there is a very strong need for 3D architectural and physical planning tools that operate in the domain of thermal, physical, and performance analysis in order to yield an optimized system implementation in 3D technology [18][39][40][41].

A few advanced CAD tools [7] have been developed however, all these tools generated by different groups, using different formats to represent the design data, create barriers for researchers who need to make use of the existing design automation tools to conduct further studies on 3D. Also, there is a lack of CAD algorithms and only few commercially EDA tools are available for 3D integrated circuits. Therefore, current 2D physical design tools such as partitioning, placement, routing, timing, extraction, must be enabled to incorporate 3D designs [42][43].

3.7 Yield and Testing Challenges

In addition to 3D design and implementation tools, there are important challenging issues in 3D test and yield that must be addressed as well. Since 3Dcircuits with multiple active layers are highly complex, it would not be economical to put off functional testing until the full process flow is complete. Instead, functionality in the different layers must be assessed along the way. Product designs and tests are challenged to provide suitable solutions.

One among the best benefits of 3D is that this technology is compatible with the known-good-die practices, a known contributor to cost reduction and test simplification.

All these design paradigms play a vital role in implementing an efficient 3D NoC architecture deriving appropriate performances suitable to the scenario. Various solutions have been proposed by several researchers to optimize these design issues and still researches are being conducted to improve the same.

Next we introduce some of the *known solutions* that have been proposed in this context like in [8] a thermal management method is proposed that uses task scheduling to limit chip temperature under required constraints. Also considers the performance degradation caused by moving task to a core far away from its data. A temperature controller is also implemented in the simulator to determine temperature management actions.

In [9] a new temperature and network competition aware mapping algorithm is proposed to reduce the peak temperature and decrease the network competition while maintaining an appropriate balance between them.

[10] Proposes a runtime distributed migration algorithm based on game theory to balance the heat dissipation among processing elements (PEs) in a 3D NoC chip multiprocessor (CMP). Modeling of this multi-objective problem as a cooperative game was possible due to the high thermal correlation between adjacent PEs in the same stack in 3D.

In [11] an efficient architecture to optimize system performance, power consumption and reliability of stacked mesh 3D NoC is presented. Stacked mesh is a feasible architecture which takes advantage of the short inter-layer wiring delays while suffering from inefficient intermediate buffers. To cope with this, an inter-layer communication mechanism is developed to enhance the buffer utilization, load balancing, and system fault-tolerance. The mechanism benefits from a congestion-aware and bus failure tolerant routing algorithm for vertical communication.

[12] Presents a power-aware run-time incremental mapping algorithm for 3-D NoCs that aims to minimize the communication power for each incoming application as well as reduce the impact of the mapped applications on future applications that are yet to be mapped. In this algorithm, if the vertical links are found to be shorter and provide higher

communication bandwidth than horizontal links, more communications can be mapped to vertical links to reduce delay and power consumption. Extensive experiments have been conducted to evaluate the performance of the proposed algorithm and the results are compared with the other heuristic algorithms.

In [13] serialization of vertical TSV interconnects in 3D ICs is proposed as one way to address these challenges which reduces the interconnect TSV footprint on each layer. This can lead to a better thermal TSV distribution resulting in lower peak temperatures, as well as more efficient core layout across multiple layers due to the reduced congestion.

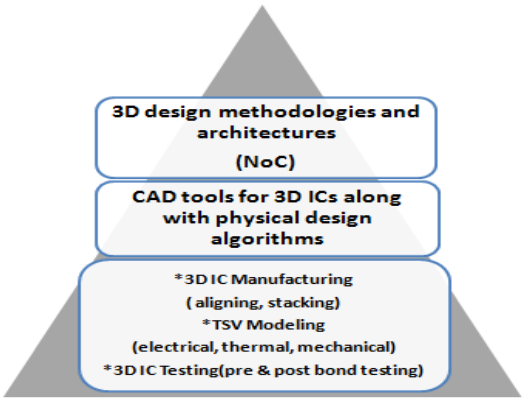

Fig 2: A concise flow of 3D IC design challenges

4. CONCLUSION

Three-dimensional NoCs are natural extensions of 2D designs. In this paper, we have tried to summarize and analyze the vast 3D IC implementation design issues explored in the current research of this domain that are beneficiary in reducing the footprint in a fabricated design along with the 3D network structures that provides a better performance compared to traditional, 2D NoC architectures.

The NoC paradigm continues to attract significant research attention both in the field of academia as well as industry. With the advent of 3D ICs, the achievable performance benefits from NoC methodology will be more renounced as shown in this paper. Consequently, this will facilitate adoption of the NoC model as a mainstream design solution for larger multicore system chips.

5. REFERENCES

[1] M. Ali, M. Welzl, and M. Zwicknagl,"Networks on Chips: scalable interconnects for future Systems on Chips," ECCSC 2008. 4th European Conference. Bucharest, pp. 240-245, July 2008.

[2] A.M. Rahmani, K. Latif, P. Liljeberg, J. Plosila, and H. Tenhunen, "Research and practices on 3D Networks-on-Chiparchitectures," NORCHIP, 2010. Tampere, pp. 1-6, November 2010.

[3] Y. Xie, "Processor architecture design using 3D integration technology," VLSI Design 2010. Banglore, pp. 446-451, January 2010.

[4] M. K. Puthal, M.S. Gaur, V. Laxmi," Performance comparison of 2D and 3D Mesh NoC," unpublished.

[5] K. Bernstein, et. al., "Interconnects in the Third Dimension: Design Challenges for 3D ICs," Design Automation Conference, 2007. San Diego, pp. 562-567, June 2007

[6] M. Ieong, et. al., "Three Dimensional CMOS Devices and Integrated Circuits," Custom Integrated Circuits Conference, 2003., in Proceedings of the IEEE 2003, pp. 207-213, September 2003.

[7] K. Salah, A. E. Rouby, R. Ragai, and Y. Ismail, "3D/TSV Enabling Technologies for SOC/NOC: Modeling and Design Challenges," Microelectronics (ICM) 2010. pp. 268-271, December 2010.

[8] H. Wang, Y. Fu, T. Liu, and J. Wang, "Thermal management via task scheduling for 3D NoC based multi-processor," SoC Design Conference (ISOCC), 2010, Seoul, pp. 440 – 444, November 2010.

[9] B. Zhang, H. Gu, Y. Yang, K. Wang, and Z. Wang, "Thermal and competition aware mapping for 3D network-on-chip," in IEICE Electronic Express, vol. 9, pp. 1510-1515, October 2012.

[10] N. Hastanpour, S. Hessabi, and P. K. Hamedani, ","N. Hassanpour, P. Khadem, S. Hessabi, "A task migration technique for temperature control in 3D NoCs," AINA 2013 in press.

[11] A. B. Ahmed, "The design of a 3D Network-on-Chip for many-core SoC," University of Aizu, October 2012.

[12] S. TYAGI," Extended balanced dimension ordered routing algorithm for 3D-networks," International Conference on Parallel Processing Workshops, pp. 499-506, 2009.

[13] T. Schonwald, J. Zimmermann, O. Bringmann, and W. Rosenstiel, "Fully adaptive fault-tolerant routing algorithm for Network-on-Chip architectures," in Digital System Design Architectures, Methods and Tools 2007, Lubeck, pp. 527 – 534, August 2007.

[14] A.M. Rahmani, et. al., "Congestion aware, fault tolerant and thermally efficientinter-layer communication scheme for Hybrid NoC-Bus 3D architectures," Networks on Chip (NoCs), 2011 Fifth IEEE/ACM International Symposium , Pittsburgh, pp. 65-72, May 2011.

[15] M. Ebrahimi, M. Daneshtalab, P. Liljeberg, J. Plosila, and H. Tenhunen, "Exploring partitioning methods for 3D Networks-on-Chip utilizing adaptive routing model," Networks on Chip (NoCs), 2011 Fifth IEEE/ACM International Symposium, Pittsburgh, pp. 65-72, May 2011.

[16] F. Li, et. al.,"Design and management of 3D Chip multiprocessors using Network-in-Memory," ISCA 2006, Boston, pp. 130 – 141, 2006.

[17] Y. Gidhini, et. al., "Topological impact on latency and throughput: 2D versus 3D NoC comparison," SBCCI, Brasilia, pp. 1 – 6, August 2012.

[18] J. Kim et al., "A novel dimensionally-decomposed router for on-chip communication in 3D architectures," in Proc. of International Symposium on Computer Architecture, pp. 138-149, 2007.

[19] B. S. Feero, and P. P. Pande, "Networks-on-Chip in a three-dimensional environment: a performance evaluation," IEEE Transactions on Computers, vol. 58, pp. 32-45, January 2009.

[20] R. I. Greenberg, and L. Guan, "An improved analytical model for Wormhole routed networks with application to Butterfly fat-trees," ," in Proc. of International

Conference on Parallel Processing, Bloomington, pp. 44-48, August 1997.

[21] C. Grecu, P. P. Pande, A. Ivanov, R. Saleh, "A scalable communication-centric SoC interconnect architecture," in Proc. of International Symposium on Quality Electronic Design, pp. 343-348, 2004.

[22] P. Guerrier and A. Greiner, "A generic architecture for on-chip packet- switched interconnections," in Proc. of Design, Automation and Test in Europe Conference, pp. 250-256, 2000.

[23] J. Kim et al., A gracefully degrading and energy-efficient modular router architecture for on-chip networks," in Proc. of 33rd International Symposium on Computer Architecture. pp. 4–15, 2006.

[24] R. Mullins, A. West, and S. Moore, "Low-latency virtual-channel routers for on-chip networks," in Proc. of 31st International Symposium on Computer Architecture. pp. 188-198, 2004.

[25] S. Heo, and K. Asanovic, "Replacing global wires with an on-chip network: a power analysis," in Proc. of the International Symposium on Low power electronics and design. pp. 369–374, 2005.

[26] R. Kumar, V. Zyuban, and Dean M. Tullsen, "Interconnections in multi-core architectures: understanding mechanisms, overheads and scaling," in Proc. of the 32nd International Symposium on Computer Architecture. pp. 408–419, 2005.

[27] Li. Shang, Li. Peh, A. Kumar, and N.K. Jha, "Thermal modeling, characterization and management of on-chip networks," in Proc. of the 37th International Symposium on Microarchitecture. pp. 67-78, 2004

[28] R. Marculescu, "Networks-On-Chip: the quest for on-chip fault-tolerant communication," in Proc. of the IEEE Computer Society Annual Symposium on VLSI,. pp. 8-12, 2003.

[29] S. Das, A. Chandrakasan, and R. Reif. "Design toolsfor 3-d integrated circuits," in Proc. of ASP-DAC. pp. 53-56, 2003.

[30] S. Alam, D. Troxel, and C. Thompson, "A comprehensive layout methodology and layout-specific circuit analysis for three- dimensional integrated circuits," in Proc. of IEEE in Symposium on Quality Electronic Design,pp. 246-251, 2002

[31] A. Rahman, and R. Reif, "Thermal analysis of three-dimensional (3-D) integated circuits (ICs)," in *Proc. of* IEEE International lnterconnectTehnology Conference. pp. 157-159, 2001

[32] R.V. Joshi, S. Kang, and C. Chuang, "3- D Thermal analysis for SO1 and its impact on circuit performance," in Proc. of SISPAD.pp. 242-245, 2001.

[33] R. S. Patti, "Three-dimensional integrated circuits and the future of system-on-chip designs," in Proc. of IEEE. vol. 94, pp. 1214 – 1224, June 2006.

[34] W. R. Davis, et. al., "Demystifying 3D ICs: the pros and cons of going vertical," in proc. of IEEE Design and Test of Computers. Vol. 22, pp. 498– 510, November 2005.

[35] B. Dang, M.S. Bakir, and J.D. Meindl, "Integrated Thermal-Fluidic I/O Interconnects for an On-Chip Microchannel Heat Sink," Electron Device Letters, Vol. 27, pp. 117-119, Feb. 2006.

[36] K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, "3-d ICs: A novel chip design for improving deep-submicrometer interconnect performance and systems-on-chip integration," in Proc. of IEEE. vol. 89, pp. 602-633, May 2001

[37] R. Weerasekera, D. Pamunuwa, M. Grange, H. Tenhunen, and L.R. Zheng, "Closed-Form Equations for Through-Silicon Via(TSV) Parasitics in 3-D Integrated Circuits," in Proc. of Workshop 3-D Integration, DATE Conference. April 2009.

[38] J. Cong,et. al., "An Automated Design Flow for 3D Microarchitecture Evaluation," in Proc. of Asia Pacific DAC. pp. 384-389, 2006.

[39] G.L.Loi, et. al., "A Thermally aware performance analysis of vertically integrated (3D) processor memory Hierarchy," in Proc. of 43rd Design Automation Conf. (DAC). pp. 991-996, 2006.

[40] O.Ozturk, F. Wang, M. Kandemir, and Y.Xie "Optimal topology exploration for application-specific 3d architectures," inProc. of Asia Pacific DAC. pp. 390-395, 2006.

[41] W.-L. Hung, G.M. Link, Y. Xie, N. Vijaykrishnan, and M. J. Irwin, "Interconnect and thermal-aware floorplanning for 3d microprocessors," International Symposium on Quality Electronic Design (ISQED). pp. 98-104, 2006

[42] C. Chiang, and S. Sinha, "The Road to 3D EDA Tool Readiness," Design and Automation Conference. pp.429-436,January 2009.

[43] A. Rahman and R. Reif. "System-level performance evaluation of three-dimensional integrated circuits," IEEE Trans. on VLSI Systems, Special Issue on System-Level Interconnect Prediction. vol. 8, pp. 671-678, 2000.

[44] International Technology Roadmap for Semiconductors http://www.itrs.net [Accessed : 31 May, 2017]

[45] [45] Wadhwani, P., Choudhary, N., & Singh, D. (2013). Energy Efficient Mapping in 3D Mesh Communication Architecture for NoC.*Global Journal of Computer Science and Technology*, *13*(14).