# TSV Multiplexing: A 3D NoC Occupancy Analysis

Yan Ghidini, Matheus Moreira, Thais Webber, Ney Calazans, César Marcon
Pontifical Catholic University of Rio Grande do Sul – Porto Alegre, Brazil
{yan.souza, matheus.moreira, thais.webber}@acad.pucrs.br, {ney.calazans, cesar.marcon}@pucrs.br

Networks-on-chip (NoCs) are proposed as promising packet-based communication platforms for multi-processor system-on-chip (MPSoC) design, due to scalability, better throughput and reduced power consumption [1]. NoC-based architectures are characterized by various trade-offs with regard to structural characteristics, performance specifications, and application demands. However, increasing the number of cores over a 2D plane is not efficient due to long network diameter and overall communication distance. In this context, with the emergence of viable 3D integration technology [2], opportunities exist for chip architecture innovations.

Indeed, 3D integration has attracted significant attention in recent years because of its potential benefits, including smaller chip footprints, higher transistor density, shorter wiring delays, and significantly higher communication bandwidth. Albeit, a wide variety of technologies are available for 3D interconnection, Through Silicon Vias (TSVs) are of particular interest for NoC-based MPSoC design [3]. Yet, the number of TSVs in a design is limited by many factors, like TSV diameter and pitch [4] and other aspects, such as power supply on 3D designs [5] and clock distribution [6].

This work analyzes network and resources occupancy for a 3D mesh NoC architecture called Lasio [7], which is an extension of Hermes 2D NoC [8]. Results demonstrate reduced links occupancy, which potentially leads to higher throughput in the NoC and more power and latency efficient systems. In addition, results suggest the possibility of TSV multiplexing without performance degradation at system level.

Lasio routers have the same types of mechanisms and resources of Hermes routers, with two additional physical ports to support 3D up/down communication, as showed in Figure 1. In this way, Lasio has five ports dedicated to connections made with routers within a same layer (Local, North, South, East and West) and two other ports (Top and Bottom) that ensure the communication between adjacent layers. Each communication port includes input and output channels, which has a buffer working as circular FIFO with configurable size. The Local port establishes a communication between the router and its corresponding PE, while the remaining ports are connected to neighboring routers. Also, Lasio topology proposes that each router has a different number of ports, depending on its position regarding to the limits of the network. More details about Lasio implementation can be found in [7].

The router architecture designed includes control logic, responsible for routing and arbitration. Lasio implements XYZ routing algorithm, which is an extension of the XY routing algorithm employed in 2D NoCs. This routing algorithm is deadlock free and requires small area for its implementation. The arbiter uses a dynamic rotating policy that prioritizes the packet routing on the input port. In other words, the arbitration is implemented using Round Robin algorithm. This method ensures that all incoming requests are processed, preventing starvation phenomenon.

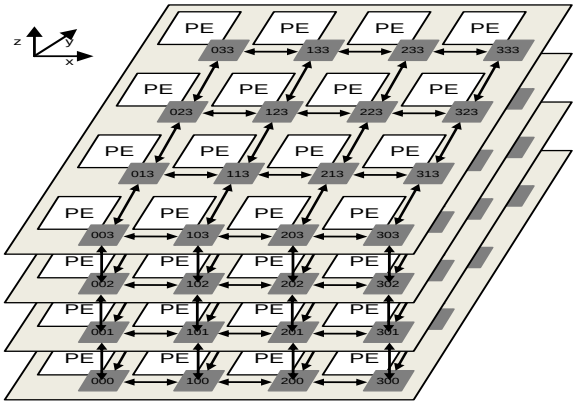


Figure 1 – The mesh-based 3D NoC architecture (4x4x4).

The parameters indicated below were used during the Lasio performance analysis and evaluation presented in this work.

1. Injection rate: packets injection rates of 100% of the maximum channels capacity was implemented and evaluated. This percentage can be translated as 800Mbps injection rate;
2. Buffer depth: 4 flit positions;
3. Synthetic traffic scenario: All-to-all – In this scenario routers send the same quantity of data (uniform packet load) in a deterministic way to all others routers, except to itself.
4. Packet size: The size used in the experiment was 8 flits (considering 16 bits width for each flit).
5. Application size: Lasio has the possibility to define the application size and also, it undertakes of determining how many packets per router are required to transmit the whole application. This number is given as (App / (Pk – 2)), where App and Pk are the application size and the packet size, respectively. The experiment realized considered an App equals to 4032 flits.

All experiments presented here assume an architecture containing 64 tiles and routers in a cubic format, 4x4x4 mesh. For this evaluation we assume that Lasio contains no virtual channel and it is credit based flow control. Moreover, inter-layer (vertical links) and intra-layer (horizontal links) hops are indistinguishable, which means the hops between

layers and the hops within the same layer have the same cost. However, we acknowledge that vertical links are normally considered a bottleneck in 3D NoC design, mainly for implying links with more area, and sometimes mechanisms to serialize and deserialize the communication.

Figure 2 shows the measured average and highest top ports buffer occupancy for the study case system. The results were obtained by measuring the average percentage occupancy for all routers of the NoC. As the charts show, higher layers present typically higher occupancy levels. One of the reasons for that is the fact that these layers will generally present higher congestion traffic scenarios, given the XYZ algorithm. However, more importantly, the obtained results suggest that vertical communication channels are typically underused. In the worst case, buffers occupancy reached a peak of 31% and an average of less than 22%. Given the 100% injection rate, and the high congestion all-to-all scenario, these values represent upper bounds for the usage of top ports. In this way, schemes for multiplexing the TSVs required by such ports are enabled.

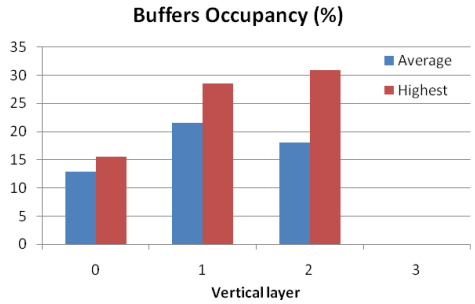


Figure 2 – Top ports buffer occupancy.

TSV sharing potentially leads to substantial area and power improvements. For instance, given the worst case peak of 31% of occupancy, the top TSVs of the network could employ 3-to-1s multiplexers. Also, preliminary results, points that TSV multiplexing will jeopardize network latency and further optimizations can be obtained by application mapping and packet, flit and buffers dimensioning. These results also display that similar occupancy levels are obtained for bottom ports buffers. In this way, we believe that all inter-layer links of the NoC can take advantage of employing TSV multiplexing.

A deeper analysis was performed on the distribution of the occupancy displayed in Figure 2. Figure 3 (a), (b) and (c) displays the occupancy of top ports buffer for each router of the three lower layers, 0, 1 and 2, respectively. Results for the top layer are omitted due to the fact that this layer has no connection to the top. As the tables suggest, further optimizations can be obtained by exploring multiplexed TSVs partitioning. For instance, given the sets of routers A= {00, 01, 11} and B= {22, 23, 32, 33}, set A typically presents higher peaks of occupancy. Its peak is of almost 31%, while the peak for set B is of less than 25%. In this way, set A should employ 3:1 TSVs multiplexing, while set B could be more relaxed and employ 4:1. Also, these optimizations can be explored at the different layers.

| Y\X | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 14.72% | 13.27% | 12.04% | 11.14% |
| 1 | 15.52% | 14.07% | 14.60% | 12.06% |
| 2 | 13.73% | 13.75% | 12.91% | 12.69% |
| 3 | 11.31% | 11.24% | 12.30% | 10.66% |

(a)

| Y\X | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 28.58% | 20.46% | 20.63% | 18.50% |
| 1 | 26.53% | 25.29% | 24.13% | 19.32% |
| 2 | 23.87% | 21.86% | 24.21% | 19.02% |
| 3 | 19.01% | 19.03% | 19.64% | 15.93% |

(b)

| Y\X | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 30.90% | 16.44% | 16.56% | 12.74% |
| 1 | 23.57% | 23.52% | 24.21% | 16.69% |
| 2 | 18.04% | 18.33% | 20.72% | 14.79% |
| 3 | 12.26% | 14.16% | 14.70% | 10.90% |

(c)

Figure 3 – Top ports buffer occupancy for each router of the 3 lower layers 0 (a), 1 (b) and 2 (c).

This work analyzed the occupancy of a study case 4x4x4 3D NoC architecture called Lasio. Results show that top ports buffers are typically underused, which allow TSV multiplexing schemes. These schemes may also be implemented at different layer levels depending on the employed routing algorithm and traffic scenario. TSV multiplexing enables higher design space exploration, in order to reduce power and area of 3D NoCs. In this way, the technique can help coping with technological issues, especially the limited number of TSVs in a 3D system.

This work is partially funded by FAPERGS PqG 12/1777-4 and Docfix (SPI n.2843-25.51/12-3). Financial support also granted by CNPQ and FAPESP to the INCT-SEC processes 573963/2008-8 and 08/57870-9.

## REFERENCES

[1] A. Jantsch and H. Tenhunen, “Network on Chip” Kluwer Academic Publishers, 2003, 312p.

[2] W. R. Davis et al. “Demystifying 3D ICs: The pros and cons of going vertical” IEEE Design & Test of Computers, 2005, pp. 498-510.

[3] International Technology Roadmap for Semiconductors. “Interconnect”, 2011 Edition. Available at http://www.itrs.net, 2012.

[4] W. K. Mak, “Rethinking the Wirelength Benefit of 3-D Integration” in: IEEE Transactions on VLSI, 2012, v.20(12), pp. 2346-2351.

[5] H. L. Chang et. al,, “A 3D IC designs partitioning algorithm with power consideration” in: ISQED, 2012, pp. 137-142.

[6] T. Y. Kim and T. Kim, “Clock Tree synthesis for TSV-based 3D IC designs” in: TODAES 2011, v.16(48).

[7] Y. G. Souza et al., “Topological Impact on Latency and Throughput: 2D versus 3D NoC Comparison” in: SBCCI, 2012.

[8] F. Moraes et. al.”HERMES: an infrastructure for low area overhead packet-switching networks on chip” in: The VLSI Journal Integration, 2004, v.38, n.1, pp. 69-93.

[9] E. Moreno et al. “Arbitration and Routing Impact on NoC Design”. In: International Symposium on Rapid System Prototyping, pp.193-198, 2011.

**Yan Ghidini** received the bachelor´s degree from the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil, in Computer Engineering in 2012, and he is currently a M.Sc. degree student in Computer Science in the same University. His research interests include intrachip communication networks, embedded system design and implementation and multi-processor systems-on-chip.

**Matheus Moreira** received his bachelor´s degree from the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil, in Computer Engineering in 2011, M.Sc. degree in Computer Science in 2012, and is currently a Ph.D. candidate in the same University. Also, he is currently an Assistant Professor at PUCRS. His research interests include asynchronous circuits design, networks-on-chip and multi-processor systems-on-chip.

**Thais Webber** received the Ph.D. degree in Computer Science from the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Brazil, in 2009. She is currently a Computer Science Post-doctoral researcher in the same university, working on hardware and software design projects and systems performance evaluation. Since 2011, she has been working on performance analysis in several areas such as modeling and simulation of parallel and distributed systems and wireless networks.

**Ney Laert Vilar Calazans** received the M.Sc. degree in Computer Science in 1988, from Federal University of Rio Grande do Sul (UFRGS), Brazil, and the Ph.D. degree in Microelectronics in 1993, from the Université Catholique de Louvain (UCL), Belgium. He is currently a Professor at the Catholic University of Rio Grande do Sul (PUCRS). His research interests include intrachip communication networks, non-synchronous circuits, computer-aided design techniques and tools. Professor Calazans is a member of the IEEE and of the Brazilian Computer Society, SBC.

**César Marcon** received the M.Sc. degree in Computer Science in 1992, from Federal University of Rio Grande do Sul (UFRGS), Brazil, and the Ph.D. degree in Computer Science in 2005, from the same University. He is currently an Associate Professor at the Catholic University of Rio Grande do Sul (PUCRS). He has been responsible for several R&D projects funded by public agencies and/or industries. His research interests include intrachip communication networks, embedded system design and implementation, and computer-aided design techniques and tools.