

Network-on-Chip, New Designs and the Path to Finding the Elusive Optimum One

Syeda Mahnur Asif
sasif@khi.iba.edu.pk

Emaan Hasan
ehasan@khi.iba.edu.pk

Abstract – The drawbacks of System on Chip, most importantly lack of scalability, have prompted new architectures, such as Network on chip (NoC) model, to come into being as a better alternative. Since first proposed, every year more and more new topologies and designs of NoC are presented in an effort to find the one with the best performance. In this paper, recent topologies that had been evaluated according to the chosen metrics of transport latency, energy dissipation and message throughput are compared and analyzed and the topologies with the best overall performance are highlighted.

Index Terms – Network on chip, topology, hybrid NoC, mesh, throughput, latency, energy.

1 INTRODUCTION

System on Chip (SoC) methodologies are currently undergoing revolutionary changes, most of which are effectively taking care of all the many drawbacks associated with them. The most potent of these changes is the Network on Chip (NoC) model that is particularly effective in tackling SoC limitations that are a result of long interconnections. These long interconnections are the main cause of non-scalable Global wire delays. NoCs, do not have these long interconnections, so provide superior latency, throughput and power statistics than their predecessors, System on Chips. NoCs employ higher levels of parallelism, use modularity to do away with the global wires, and utilize locality for power efficiency. NoCs further enable widespread integration of Intellectual Property (IP) cores into singular SoCs. Owing to these immense benefits of the NoC model, a large amount of research in recent years has been directed towards making this model as efficient as possible. This has largely been done by exploring multiple architectural topologies with the aim of finding the most suitable in terms of reducing latency, energy dissipation, area overhead and increasing throughput. The topology is what determines the number of hops, and the interconnect lengths between source, destination and intermediary routers. The amount of links and types of wires are important factors in terms of complexity costs of the chip and

these along with the aforementioned factors are all dependent on the NoC topologies.

This paper explores the topologies at the forefront of the race to finding the best design. We will first define the common metrics that will be used to compare the NoCs. The section following after will outline the topologies selected, and the next section will then highlight the drawbacks and advantages of each.

2 COMPARISON METRICS

In order to effectively compare and contrast various NoC architectures, a set of parameters must be established that give an accurate picture of one architecture's superiority over another [1]. The metrics we make use of in this paper are as follows.

2.1 Transport Latency

Defined as the amount of clock cycles it takes between a message header to be sent from the source node, and that message's tail, to exit the destination node. The message travels in flits. Flits are fixed length flow control units. Latency consists of the time needed for the flits to travel the path consisting of switches and interconnects as well as some overhead caused by

sender and receiver. Therefore, for overall latency L is given by:

$L = \text{sender overhead} + \text{transport latency} + \text{receiver overhead}$.

2.2 Energy dissipation

Due to the logic gates and the inter-switch wires toggling, energy is dissipated when a flit travels in an NoC. Energy is dissipated at each hop that the message takes when traveling between the source and destination. Each hop consists of wires and switches. The energy dissipated at each of these hops depends on the total capacitance and signal activity of the switch or the interconnect wire used. So our parameter is the energy dissipated when transporting a packet consisting of n flits over h hops.

2.3 Message throughput

This is the rate at which message traffic is being sent across the NoC. Throughput is given by the product of total messages completed and message length, divided by the number of IP blocks and Total time. This makes message throughput, a fraction of the maximum load that the NoC is capable of handling.

3 TOPOLOGIES UNDER DISCUSSION

3.1 Star Topology

As mentioned by work done at the *Department of Electrical Engineering, Department of Computer Science and Information Engineering, Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University* [2], the Star topology is an $N \times N$ 3D representation of the standard 2D mesh topology converted into proposed 3×3 sub meshes. All the sub meshes have their central nodes connected to each other's central and diagonal nodes. The nodes mentioned here are routers or switches. Eight additional ports are added to the initial 5-port router for connecting the two levels of meshes. These ports are put in place to make connections along the direction of the second-level mesh and with the diagonal router in the first level mesh.

3.2 Mesh-Ring Topology

As proposed by [3] the Mesh-Ring topology is where the network is divided into "subnets," essentially clusters of neighboring cores. This topology is a 2×2

mesh subnet that uses the switches and links of a standard mesh topology. A single central hub connects all the links and hubs from the clusters, forming an entire network level of subnets. This generates a 32 IP block.

3.3 Mesh-Torus-folded Torus Hybrid Topology

As proposed by [4], this topology consists of a 4×4 hybrid connected via torus links, folded tori links and mesh links.

a) *Torus links*: The links connected to all the boundary routers are the same as that of torus NoC topology which is creating a wrap around connection to reduce the hop count.

b) *Folded torus links*: The links which are connecting the even and odd nodes together like a folded torus topology.

c) *Mesh link*: These links are utilized to connect the adjacent routers other than the diagonal routers.

The boundary routers except the diagonal edge router utilize all three kinds of links (one torus link, 2 folded torus links and a single hop link). All inner routers utilize 2 single hop links and 2 folded torus links.

3.4 A novel busmesh NoC

The work proposed in [5] utilizes packet transmission priority control method. It is composed of cluster nodes (CNs) and mesh routers (MRs). In intra-cluster node, several cores which have heavy communication to each other are connected by a local bus. The suggested architecture is a generalized and simplified version of hybrid NoC.

a) *Hybrid NoC*: This is a network on chip architecture that has local buses and global mesh routers.

b) *Busmesh NoC (BMNoC)*: A NoC architecture with clusters which are connected by mesh network, borrowing the hierarchical model from the Internet and adapting it to communication networks.

3.5 NoC based on partial interconnection of mesh networks

A new architecture proposed in [6] is based on partially connected mesh topology. Four extra bidirectional channels are added to each router of the mesh network. This forms nine bidirectional communication channels in each router.

a) *Partially connected mesh*: This is a network topology in which some of the nodes of the network are connected to more than one other node in the network with a point-to-point link.

3.6 ARB-NET-based 3D Hybrid NoC-Bus mesh architecture

An integrated low-cost monitoring platform for 3D stacked mesh architectures is proposed in [7] which can be efficiently used for various system management purposes such as traffic monitoring, thermal management and fault tolerance. The proposed infrastructure called ARB-NET utilizes bus arbiters to exchange the monitoring information directly with each other without using the data network.

a) *3D stacked mesh*: This topology integrates multiple layers of 2D Mesh networks by connecting them with a bus spanning the entire vertical distance of the chip [8]. For communications, it uses a hybrid between a packet-switched network and a bus.

4 SUMMARIZED RESULTS

In this section the experimental results of the aforementioned topologies as appeared in their respective papers will be referenced and summarized so they can be used in the later section for analysis.

The star-type NoC performed well for smaller networks, showing improvements of 17.3% for a 12×12 mesh, 3.85% for a 6×6 mesh and 12.90% for a 9×9 mesh. Moreover, when the network size increased, the performance improved without excessive power and area overheads unlike the normal NoC. Thus, improvements of 19.76% for a 15×15 mesh and 21.43% for an 18×18 mesh could be obtained.

The experimental results in [2] showed that in comparison with the normal mesh NoC, the star-type NoC consumed 34.27% more power and 57.54% more area, while reducing the hop count by 60.87% and latency is reduced by 155%. Furthermore, in comparison with the level-2 mesh NOC, the power overhead of the star-type NOC was 1.86% and the area overhead, 10.77%, while the hop count reduction was 18.18%. Finally, in comparison with the overall performances (considering power, area, and latency) of the normal and level-2 mesh NoCs, the overall performance of the star-type NoC showed improvements of 17.3% and 10.28%, respectively.

The hybrid mesh-ring design proposed in [3] had its performance evaluated against its wired counterpart. Results showed that throughput of the hybrid model increased by 31%. After calculations regarding power dissipation in both designs, it was found that the wired design dissipated 2.3 W while the wireless design dissipated 2.06 W which proved that the hybrid wireless design power dissipation was better than wired model

by 11%. Latency was improved by 20% in the hybrid model.

The performance of the topology in [4] was compared to the topologies it was composed of, namely, mesh, torus and folded torus topologies. The authors found that the ideal throughput of hybrid topology increases by 200% when compared to mesh, 50% compared to that of torus and folded torus, the 'Hmin' is decreased by 32% when compared to mesh, 5% compared to torus and folded torus topologies and the latency is decreased by 26% when compared to mesh, 4% compared to torus and folded torus topologies. In terms of power consumption, the proposed hybrid consumed 38% less than mesh, 7% less than torus and 8% less than folded torus topology.

The experimental evaluations of the work proposed in [5] showed that BMNoC utilizing packet transmission priority control method improved the critical traffic load by approximately 20% as compared to Hybrid NoC (HNoC) and approximately 15% as compared to HNoC using packet transmission priority control method (PTPCM). BMNoC+PTPCM improved the critical traffic load as compared to conventional BMNoC (approximately 6% improvement). At low traffic loads, the average packet latency exhibits a weak dependence on the traffic injection rate. However, when the traffic injection rate exceeds a critical traffic load, the packet delivery cycles rise abruptly and the network throughput starts collapsing [5]. The average packet latency for BMNoC+PTPCM was consistently smaller as compared to HNoC, HNoC+PTPCM and conventional BMNoC. No data was provided by the authors regarding the energy efficiency of the proposed design.

The proposed architecture and routing algorithm in [6] were compared to measure performance benefits over standard mesh network in terms of delay and throughput. Significant improvement in average latency (60% reduction) and overall average throughput (60% increased) were observed when using the proposed network. However, an increase in number of channels made the switches expensive and could increase the area and power consumption.

In [7] the proposed ARB-NET based architecture using the *AdaptiveXYZ* routing algorithm showed 19%, 9%, and 4% drop in power consumption over the Symmetric 3D-mesh NoC, typical 3D NoC-Bus Hybrid mesh, and *AdaptiveZ* 3D NoC-Bus Hybrid mesh architectures, respectively. Similarly, 29%, 17%, and 10% reduction in average packet latency over the Symmetric 3D-mesh NoC, typical 3D NoC-Bus Hybrid mesh, and *AdaptiveZ* 3D NoC-Bus Hybrid mesh architectures was also observed for the proposed architecture. The message

throughput generally increased as bus utilization was more efficient which made the load, balanced.

5 COMPARISON AND ANALYSIS

From the experimental results in the respective papers, also summarized above, it can be observed that [3] and [5] compare the performance of their proposed works against the typical 2D Hybrid NoC, whereas [7] used 3D Hybrid NoC amongst other topologies, as its comparison. Since three-dimensional NoCs are natural extensions of 2D designs [8] they can be duly compared.

Power consumption decreased by 11% in [3] and by 9% in [7]. No information was provided by the authors in [5] with regards to this comparison metric. Throughput increase by 31% in [3] and there was a general increase in [5] and [7]. Transport latency improved by 20% and 17% in [3] and [7] respectively, while [5] concluded a general improvement over the typical Hybrid NoC.

The experimental results in [2], [4] and [6] compare the performance of their proposed works against the typical 2D mesh.

With regards to power consumption, a 34.27% increase, 38% decrease and an inferred general increase in observed respectively. Throughput is increased by 200% in [4] and 60% in [6]. The experimental results in [2] do not quantitatively quote a numerical figure for the message throughput of their design. However, a value for the minimum hop count is given. H_{min} is observed to decrease by 60.87%. Since, the average hop count has a definitive effect on throughput [8] thus, it can be deduced that throughput of [2] does increase with regards to NoC mesh, although more data is required to calculate the exact numbers. Latency in [2], [4] and [6] is reduced by 155%, 26% and 60% respectively.

6 CONCLUSIONS AND FUTURE WORK

From the comparison of performances in the previous section it can be observed that against a typical 2D mesh architecture, the proposed Mesh-Torus-folded Torus Hybrid Topology in [4] has the best energy efficiency and message throughput, it consumed 38% less power and increased throughput by 200%, while the star topology in [2] reduced the transport latency the most, by 155%.

Compared against the Hybrid NoC, the work in [5] is disregarded due to the lack of data for our comparison metrics, while Mesh-Ring topology in [3] is concluded to have the best overall performance, with reductions in

power consumption by 11%, in latency by 20%, and an increase in throughput by 31%.

Since the authors in [8] conclude that besides reducing the footprint in a fabricated design, 3D network structures provide a better performance compared to traditional, 2D NoC architectures, future work should be done in integrating the topologies highlighted as best amongst the ones discussed, into 3D architectures.

REFERENCES

- [1] P. Partha Pratim, *et al.*, "Performance evaluation and design trade-offs for network-on-chip interconnect architectures," *Computers, IEEE Transactions on*, vol. 54, pp. 1025-1040, 2005.
- [2] Kuan-Ju Chen; Chin-Hung Peng; Feipei Lai, "Star Type Architecture with Low transmission Latency for a 2D Mesh NOC", *Circuits and Systems (APCCAS)*, 2010 IEEE Asia Pacific Conference, 2010, pp. 919 – 922.
- [3] Abd El Ghany, M.A.; et al., "Hybrid Mesh-Ring wireless Network on Chip for multi-core system," *ISOC*, 2012.
- [4] Swaminathan, K.; et al, "A Novel Hybrid Topology for Network on Chip", *IEEE 27th Canadian Conference on Electrical and Computer Engineering*, pp. 1–6, 2014.
- [5] Seungju Lee, Nozomu Togawa, Yusuke Sekihara, Takashi Aoki, and Akira Onozawa, "A hybrid NoC architecture utilizing packet transmission priority control method," in *Circuits and Systems (APCCAS)*, 2012 IEEE Asia Pacific Conference on, 2012, pp. 404-407.
- [6] Choudhary, S.; Qureshi, S., A new NoC architecture based on partial interconnection of mesh networks, 2011 IEEE Symposium on Computers & Informatics (ISCI), pp.334,339, 20-23 March 2011 doi:10.1109/ISCI.2011.5958937
- [7] A.-M. Rahmani, K.R. Vaddina, K. Latif, P. Liljeberg, J. Plosila, and H. Tenhunen. Generic Monitoring and Management Infrastructure for 3D NoC-Bus Hybrid Architectures. In *Proceedings of the 6th ACM/IEEE International Symposium on Networks-on-Chip*, pages 177–184, 2012.
- [8] B.S. Feero and P.P. Pande. Networks-on-Chip in a Three Dimensional Environment: A Performance Evaluation. *IEEE Transactions on Computers*, 58(1):32–45, 2009.