Network-on-Chip: A New SoC Communication Infrastructure Paradigm

Naveen Choudhary

***Abstract*— As the feature size in deep-submicron domain is continuously shrinking and the bandwidth requirements is increasing, traditional shared-bus architecture will no longer be able to meet the requirements of System-on-Chip (SoC) implementations. Specially, inherently non-scalable nature of the shared-bus architecture as well as its power hungry nature will become the communication bottleneck in most practical applications. Network-on-Chip (NoC) communication architectures have emerged as a promising alternative to address the problems associated with on-chip buses by employing a packet-based micro-network for inter-IP communication. Some of the most important phases in designing the NoC are the design of the topology or structure of the network and setting of various design parameters (such as frequency of operation, link-width, etc). This paper surveys the various topological structures for NoC proposed in the research domain**

***Index Terms*—NoC. SoC, Topology, Routing, Buffers, Virtual Channel.**

I. INTRODUCTION

With the advances in the Deep Sub-Micron (DSM) technology, the huge number of transistors available on a single chip allows designers to integrate tens or hundreds of IP blocks together with large amounts of embedded memory. These IPs can be CPU or DSP cores, video streaming processors, high-bandwidth I/O, etc [1].

This richness of the computational resources places tremendous demands on the communication resources as well. Additionally, the shrinking feature size in the deep-submicron (DSM) domain makes interconnect delay and power consumption the dominant factors in the optimization of modern systems. Another consequence of the DSM is the dificulty in optimizing the interconnect due to the worsening effects such as crosstalk, electro-magnetic interference and soft errors, etc [2]. To date, the shared-bus scheme (either single bus or multi-bus) has been the system communication architecture of choice. However, there are several problems associated with the standard bus architectures. First, a global bus implies a large capacity load for the bus drivers. In turn, this implies large delays and huge power consumption. Second, the performance of shared bus architecture is inherently not scalable as there can be at most one transaction over the shared bus at any point of time. Moreover, the bus performance has to be degraded if a slow device is accessing the bus. To address this problem, some sophisticated modern bus architectures address this problem through the concept of bus hierarchy and separation. However, such a temporary solution falls short when hundreds or over thousands of processors will have to be integrated on a single chip in the near future [2]. Third, in DSM era, design of long, wide buses becomes a real challenge. While physical information is extremely important for successful bus design, the environment in which the bus is embedded is very hard to predict and characterize early in the design stages. The scalability and success of switch-based networks and packet-based communication in parallel computing and Internet has inspired the researchers to propose the Network-on-Chip (NoC) architecture as a viable solution to the complex on-chip communication problems [3]. In Section 2 popular regular NoC architectures are discussed. Section 3 discusses the usefulness of Irregular topology for the NoC. In Section 4 recently proposed 3D-NoC architecture is discussed. Section 5 presents some simulation experimental results for regular and irregular NoCs and in Section 6 we conclude.

Dr. Naveen Choudhary, Department of Computer Science and Engineering, College of Technology and Engineering, MPUAT, Udaipur, India. (e-mail: naveenc121@yahoo.com).

II. REGULAR NETWORK-ON-CHIP

The topology of a NoC specifies the physical organization of the interconnection network. It defines how nodes, switches and links are connected to each other. Topology for NoCs can be classified into two broad categories: 1) direct network topologies, in which each node (switch) is connected to at least one core (IP/PE), and 2) indirect network topologies, in which we have a subset of switches (nodes) not connected to any core (IP/PE) and performing only network operation. Both direct and indirect topology can be regular like meshes, tori, k-ary n-cubes and fat trees or irregular topology. Some examples of regular topologies are shown in Figure 1. Most NoCs implement regular forms of network topology that can be laid out on a chip surface (a 2-dimensional plane) for example, k-ary 2-cube (where k is the degree of each dimension and 2 is the number of dimensions) commonly known as grid-based topologies. Besides the form, the nature of links adds an additional aspect to the topology. In k-ary 2-cube networks, popular NoC topologies based on the nature of link are the mesh which uses bidirectional links and torus which uses unidirectional links. For a torus, a folding can be employed to reduce long wires. In the NOSTRUM NoC presented by Millberg et al. [4], a folded torus is discarded in favor of a mesh with the argument that it has longer delays between routing nodes. Generally, mesh topology makes

better use of links (utilization), while tree-based topologies are useful for exploiting locality of traffic. The standard regular topologies such as meshes, tori, k-ary n-cubes or fat trees as shown in Figure 1 are popularly used as the wires can be well structured in such topologies. NOSTRUM [5], SOCBUS [6] are regular 2D-Mesh architectures. In [7] torus architecture (NTNU) is described. An alternate FAT tree based structure is used in the SPIN [8] and PROPHID [9] approaches. These approaches are adequate for general purpose systems where the traffic characteristics of the system cannot be predicted statically, as in homogeneous chip-multiprocessors [10].

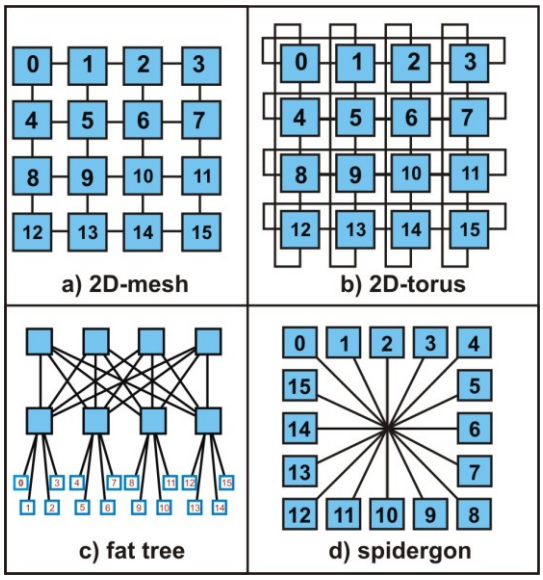

Fig 1: Regular topology based Network-on-Chip

The *k-ary tree* and the *k-ary n-dimensional* fat tree are two popular regular topologies for regular NoC. Where *k* is the degree of each dimension and *n* is the number of dimensions. The network area and power consumption scales predictably for increasing size of regular NoCs. Most NoCs prefer regular topologies that can be implemented on a 2-dimensional plane of a chip. Such topologies are generally referred as grid-based topologies. Another popular regular NoC is Octagon NoC [11]. Its basic configuration is a ring of 8 nodes connected by 12 bidirectional links which provides two-hop communication between any pair of nodes in the ring and a simple shortest-path routing algorithm can be used for packet routing. Such rings can then be connected edge-to-edge to form a larger, scalable network.

The XY routing [12, 13] and odd-even routing [14] are the most used deadlock free routing algorithms for the popular 2D-mesh based NoCs. They are both theoretically guaranteed to be free of deadlock and livelock. The XY routing strategy can be applied to regular two-dimensional mesh topologies without obstacles. The position of the mesh nodes and their nested network components is described by coordinates, the x-coordinate for the horizontal and the y-coordinate for the vertical position. A packet is routed to the correct horizontal position first and then in vertical direction. XY routing produces minimal paths without redundancy, assuming that the network description of a mesh node does not define redundancy. The odd-even turn model is a shortest path routing algorithm that restricts the locations where some types of turns can take place such that the algorithm remains deadlock-free. More precisely, the odd-even routing prohibits the east to north and east to south turns at any tiles located in an even column. It also prohibits the north to west and south to west turns at any tiles located in an odd column.

III. IRREGULAR NOC

Many high end complex application specific SoCs are heterogeneous in nature, with each core having different size, functionality and communication requirements. For such application specific irregular NoC it is desired to have a customized irregular topology. Figure 2 exhibits an example irregular topology.

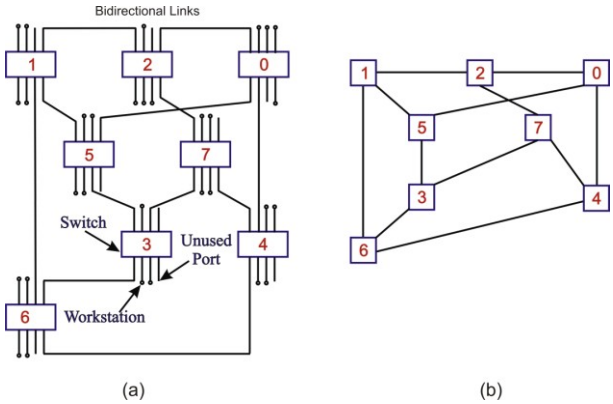

Fig 2: An example application specific NoC with irregular topology having asymmetric structure (a) irregular NoC (b) corresponding topology graph

In addition to above mentioned applicability of irregular topologies for application specific NoCs, the generic regular topologies can also become irregular for supporting oversized region or due to faults in switches or links in the regular NoCs. The region concept presented in [15] was intended for use of larger resources which do not fit in the fixed sized slot of a regular mesh architecture layout as shown in figure 3. Nevertheless, this concept can be used in a variety of other contexts [16] as mentioned below.

- Encapsulating a group of resources with special requirements on performance, power consumption or data security. Such a region could have specialized interconnections as well as communication protocols.
- Region as a logical structure. In this case the internal hardware design of the region is identical with the outside NoC structure. This assumes configurable routers in the NoC for defining, isolating and maintaining a region.
- Support for different configurations of power/ performance modes of resources inside a region by control of operating voltage, clock frequency etc.
- Reuse of multi-core subsystems. These solutions are currently available as separate SoCs. The concept of region offers the possibility of raising the level of reuse from a core to a level where specially designed multi-core subsystems can be reused. Without the region concept these subsystems will need to be redesigned, keeping in view the NoC constraints.

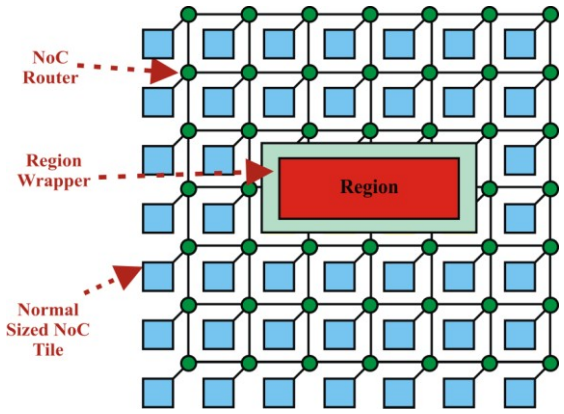

Fig 3: Mesh topology with over sized region

Moreover a regular topology of NoC can become irregular due to faults at design time. Figure 4 shows a regular 2D-Mesh topology with faults making the overall topology of the NoC as irregular.

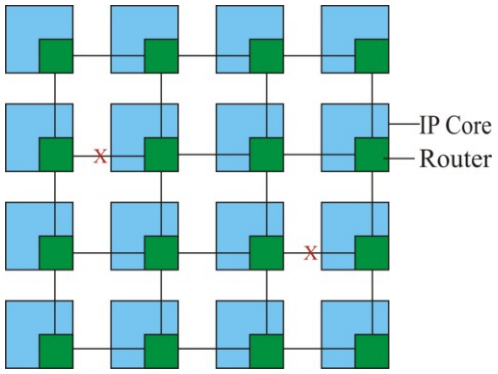

Fig 4: Irregular mesh NoC due to Faults

If the topology is regular, it is wise with regards to performance, to use a topology dependent routing since it would be able to exploit the regularity of the topology. Dimension-Order Routing (DOR) [13] is such an algorithm suitable for meshes. However for irregular NoC topology independent routing is preferred. As deadlock-free routing is critical for proper operation of such irregular NoCs, several high performance topology agnostic routing algorithms exist, such as up*/down* [17], lturn [18], down/up [19], prefix-routing [20], etc. These algorithms have in common that they are based on turn prohibition, a methodology which avoids deadlock by prohibiting a subset of all turns in the network. A problem with this approach is that it is unable to guarantee shortest path routing and unable to exploit any regularity in the underlying topology.

IV. 3D NoC

The semiconductor industry is now moving towards the 3D stacking technology for the forthcoming nano scale generation with the advancement in transistor packing density [21]. The greatest advantage for 3D NoC is that it can greatly help in reducing the diameter of the topology of NoC leading to reduction in packet transfer time and latency. Moreover the increasing the adaptivity for information routing can also help in achieving the desired throughput. A 3D communication architecture for the NoC can be established by Large portioning a 2D die into smaller segments and stacking them. The decrease in diameter of the topology for 3D architecture has a great advantage of reducing the wiring or channel on the chip. Figure 5 exhibits a example of 4x4x4 3D-Mesh based NoC.

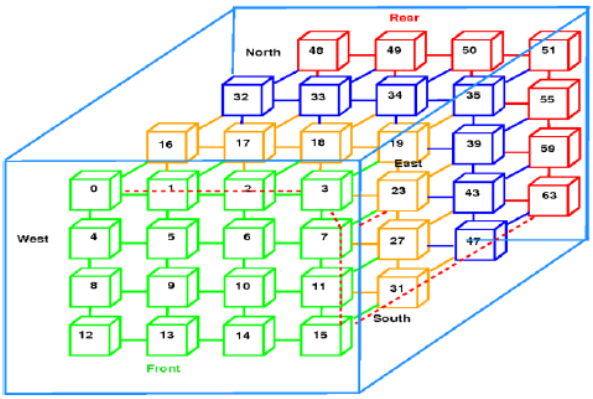

Fig 5: 4x4x4 3D-Mesh based NoC

V. Experimental Results

For performance comparison of Irregular and regular NoC, the NoC simulator *IrNIRGAM* [22] which is an extended version of NIRGAM [23] was used.

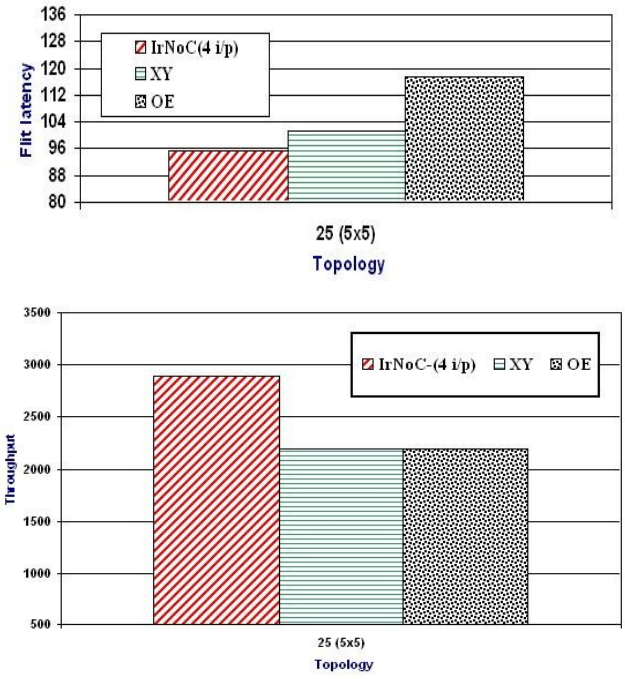

Fig 6: Performance comparison of irregular and regular (2D Mesh) NoC on the NoC simulation framework

The simulation was run for 10000 clock cycles and network throughput in flits and average flit latency were used as parameters for comparison. Network throughput is the number of flits received by various cores of the NoC during the simulation run. The flit latency determines the number of clock cycles it takes from entering the network until the reception at the target node. All data queues in the NoC for regular as well as irregular topologies was assumed to be eight

flits per channel. For regular NoC XY and OE routing were used for information transmission, whereas for irregular NoC up*/down* routing was used. The traffic pattern was taken according a multimedia application as proposed in Hu et al. [24]. The performance comparison of irregular NoC wirh regular NoC in terms of throughput and flit latency is exhibited in figure 6.

Figure 6 shows that the Irregular NoC shows better performance in terms of throughput and flit latency. The irregular NoC shows an increased throughput by 31.7% and decrease in average flit latency of 5.9 clock cycles and 22.4 clock cycles in comparison to regular NoC with XY and OE routing.

VI. Conclusion

NoC research includes a broad range of research such as communication layers, topology, switching, routing etc. NoC is relatively new area of research where various open issues of this domain are being tried to be addressed by the researchers. NoC design is a significative direction of future on-chip systems development and the hardware development of NoC architectures is growing quite rapidly. In this paper we tried to survey the various proposed topological structures in the NoC research domain.

References

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

[2] International Technical Roadmap for Semiconductors website. [online] Available: http://public.itrs.net/.

[3] W. J. Dally, B. Towles, "Route packets, not wires: on-chip interconnection networks", in Proceedings DAC, pp. 684-689, June 2001.

[4] M. Millnerg, E. Nilsson, R. Thid, A. Jantsch, "Guaranteed bandwidth using looped containers in temporally disjoint networks within the nostrum network-on-chip," in Proceedings of Design, Automation and Testing in Europe Conference (DATE). IEEE, pp. 890–895, 2004.

[5] S. Kumar, A. Jantsch, J.-P. Soininen, M. Forsell, M. Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, "A network on chip architecture and design methodology," in Proceedings of VLSI Annual Symposium (ISVLSI 2002), pp. 105–112, 2002.

[6] D. Wiklund and D. Liu, "SOCBUS: switched network on chip for hard real time embedded systems," in Proceedings of the Int. Parallel and Distributed Processing Symposium, Apr.2003.

[7] L. Natvig, "High-level architectural simulation of the torus routing chip," in Proceedings of the International Verilog HDL Conference, California, pp. 48–55, Mar. 1997.

[8] L. Benini, G. DeMicheli., "Networks on Chips: a new SoC paradigm," IEEE Computer Vol. 35, No. 1 pp. 70–78, January 2002.

[9] P. Guerrier, A. Greiner, "A generic architecture for on-chip packet-switched interconnections," in Proceedings of the Automation and Test in Europe Conference and Exhibition, pp. 250–256, 2000.

[10] M. Taylor, W. Lee, S. Amarasinghe, A. Agarwal, "Scalar operand networks," in IEEE Transactions on Parallel and Distributed Systems, vol. 16, no. 2, pp. 145-162, Feb 2005.

[11] F. Karim et al., "An interconnect architecture for networking systems on chips", *IEEE Micro 22*, pp. 36–45, 2002.

[12] W. Zhang, L. Hou, J. Wang, S. Geng, W. Wu, "Comparison research between XY and odd-even routing algorithm of a 2-dimension 3\times3 mesh topology network-on-chip," in GCIS'09, pp. 329–333, 2009.

[13] J. Duato, S. Yalamanchili, L. Ni, Interconnection Networks : An Engineering Approach, Elsevier, 2003.

[14] G. M. Chiu, "The odd-even turn model for adaptive routing," in IEEE Transactions on Parallel and Distributed Systems, vol. 11, no. 7, pp. 729–738, Jul 2000.

[15] S. Kumar, A. Jantsch, J.P. Soininen, M. Forsell, M. Millberg, J. Öberg, K. Tiensyrjä, A. Hemani, "A Network on Chip Architecture and Design Methodology", In IEEE Annual Symposium on VLSI, April 2002.

[16] R. Holsmark, S. Kumar, "Design Issues and Performance Evaluation of Mesh NoC with Regions", In IEEE NorChip, Oulu, Finland, pp. 40-43, Nov. 2005.

[17] e. a. M. D. Schroeder, "Autonet: A High-Speed Self-Configuring Local Area Network Using Point-to-Point Links", In Journal on Selected Areas in Communications, vol. 9, Oct. 1991.

[18] A. Jouraku, A. Funahashi, H. Amano, M. Koibuchi, "L-turn routing: An Adaptive Routing in Irregular Networks", In the International Conference on Parallel Processing, pp. 374-383, Sep. 2001.

[19] Y.M. Sun, C.H. Yang, Y.C Chung, T.Y. Hang, "An Efficient Deadlock-Free Tree-Based Routing Algorithm for Irregular Wormhole-Routed Networks Based on Turn Model", In International Conference on Parallel Processing, vol. 1, pp. 343-352, Aug. 2004.

[20] J. Wu, L. Sheng, "Deadlock-Free Routing in Irregular Networks Using Prefix Routing", ***DIMACS Tech. Rep.*** *99-19,* Apr. 1999.

[21] B. Black and et al., Die stacking (3d) microarchitecture. In *MICRO 39*, pages 469–479, 2006.

[22] N. Choudhary, M. Gaur, V. Laxmi, "Irregular NoC simulation Framework: IrNIRGAM", in proceeding of the international conference ETNCC 2011, pp. 1-5, April 2011.

[23] L. Jain, B. M. Al-Hashimi, M. S. Gaur, V. Laxmi, A. Narayanan, "NIRGAM: A Simulator for NoC Interconnect Routing and Application Modelling". DATE 2007, 2007.

[24] J. Hu, R. Marculescu, "Energy-Aware Mapping for Tile-based NOC Architectures Under Performance Constraints". *In ASP-DAC* 2003, Jan 2003.

Dr. Naveen Choudhary received his B.E, M.Tech and PhD degree in Computer Science & Engineering. He completed his M.Tech from Indian Institute of Technology, guwahati, India and PhD from Malviya National Institute of technology, Jaipur, India in 2002 and 2011 respectively. Currently he is working as Associate Professor and Head, department of Computer Science and Engineering, College of Technology and Engineering, Maharana Pratap University of Agriculture and Technology, Udaipur, India.

His research interest includes Interconnection Networks, Network on Chip, Distributed System and Information Security. He is a life member The Indian Society of Technical Education, Computer Society of India and The Institution of Engineers, India. E-mail: naveenc121@yahoo.com