Evolution & Performance study of 2D NoC Topologies

Sapna Tyagi, Amit Agarwal, Vinay Avasthi, Piyush Maheshwari

Abstract*: On chip architectures are adopted as communication infrastructure for System on chip. Various IP cores are integrated on planar chip because of that optimal utilization of resources and reusability can be achieved. The direct link interconnections and shared bus interconnections of SoCs are unable to meet the desired scalability, reliability, and high throughput requirements. The key design considerations and efficiency of NoC as communication infrastructure are dependent on topologies, routing algorithms , low power consumptions and optimum buffer utilization .In our work , the NoC topologies like Torus , Mesh, C-Mesh and Fattree are analyzed . The Parameters used are latency, throughput and hop-count. The comparison presented will help in understanding the topologies empirically. Results presented in the work proved that the torus topology is optimized design and exhibits good trade-off between performances and scalability among the four basic NoC architectures.

***Index Terms*: Booksim 2.0 , NoC, system on chip , Topologies**

I. INTRODUCTION

System-on-Chip (SoC) is a technology used by VLSI designers for designing today's integrated circuit (IC) chips that puts IP cores onto a single silicon die where IP cores communicate with each other and also have interface to other peripheral devices through some interconnections. To meet the escalating application and computation intensive system, the number of IP cores integrating on chip are increasing considerably every year.The major challenge here is that as number of components continue to increase the communication architecture becomes the bottleneck. Traditionally bus based system and Point to Point links based interconnection architectures can be used for the system with lesser number of IP cores and they could not scale well with this increasing number of components. There is lack of reusability and flexibility in the system. Thus the design of efficient & performance based communication infrastructure is gaining equal importance [8-12].

A solution that satisfies the need of this efficient Communication infrastructure requirement is Network-on-Chip (NoC).This has emerged as communication infrastructure proposed by Dally and Towles where various number of functional and storage cores / sub systems also known as IP Core like MAC Ethernet or PCI bus controllers, memory controllers, processors, peripheral devices, video processors and audio hardware, buffer memories, I/O peripherals, hardware accelerators, etc are fused onto a planar chip to exchange data and commands with each other by using packet switching technology on a hop-by-hop basis[1]. Network–on-Chip approach is emerged as scalable solution as compare to traditional bus-based and point-to-point communication infrastructure which eliminate point to point ad-hoc global wiring and introduced on chip packet based interconnection network borrowed from the concepts of networks. NoCs presents better scalability & reusability with more structured and modular design which can give higher bandwidth and low latency [8][9].

The major difference between Network-on- Chip approach and traditional bus-based approach are as follows:

- In a bus, the IP Cores are interconnected using wires and are combined with an arbiter to manage the access to the bus. Under NoC technology, IP Cores are interconnected using switches or routers, synchronizers, width converters, power isolator cells with user-defined topologies.
- There is scalability issue under bus interconnect , for every IP core added , parasitic capacitance increases , which degrades the electrical performance .Under NoC each IP core is linked in point-to-point manner in any topology design adopted. Due to which local performance do not degrade at the time of scaling .
- There is limited Bandwidth which is shared by all components attached in the bus based interconnects , whereas in NoC based interconnects , as with each IP core, the bandwidth scales up

Network-on-chip systematically arranges IP cores, where each IP core is accompanied by router/switch forming a Network which all together exchange information in the form of data packets through dedicated links. There are various simulators for studying and exploration of NoC like Access Noxim, Nirgam, Orion 3.0, Booksim 2.0, NoCSEP, Nostrum and JAVA NoC simulator. We have used Booksim 2.0[2] network simulator for our work. Booksim is flexible NoC simulator written in C language and presents detailed representation of NoC model. It has the capability of evaluating the topologies like torus, mesh, fat tree & C-mesh and routing algorithm like dimension order routing . The parameters used in the simulator are injection rate, throughput & latency.

The rest of the paper is structured in the following way: Section II provides Network on chip essentials for the paper, section III presents various popular NoC topologies like mesh, torus, c-mesh, fattree,

Revised Manuscript Received on December 22, 2018.

Sapna Tyagi, IT , IMT, Ghaziabad, India
Amit Agarwal, School of Computer science University of Petroleum & Energy Studies, Dehradun, India
Vinay Avasthi, School of Computer science University of Petroleum & Energy Studies, Dehradun, India
Piyush Maheshwari, Department of Engineering, Amity University Dubai United Arab Emirates

section IV presents simulation of various NoC Topologies and evaluated them in terms of injection rate, throughput and latency using Booksim 2.0 and finally the conclusion part is described in section V.

II. NETWORK-ON –CHIP ESSENTIALS

The fundamental components of NoC are Network Interface, Routing algorithms, Topologies & flow control mechanism and router microarchitecture.

1. Topology: NoC topology is important and key design factor in NoC. It has huge implication on the performance and cost and should be regular and simple so that simple and efficient routing can be implemented. Network topology specifies as the how various processing elements (IP's, peripheral devices, controllers, DSP's, and Processors etc.) are physically organized on single chip. Topologies can be well understood with the help of indirected graph G(R, N, and L) where R presents all Routers and N is the set of IP cores and L links. Topologies can also be categorized into Direct and Indirect. In the direct Topology each node has associated router which further connected to fixed number of neighboring nodes whereas in the Indirect topologies some routers/switches do not have processing element (PE), instead they are only dedicated to perform routing functions and flow control for entire network.

2. Routing Algorithms: They are the key design consideration for overall performance of NoC. Algorithm is responsible for determining the path taken by a packet between source and destination nodes on a given topology. The main task of the algorithm is to :

- Prevent livelock ,deadlock, and starvation conditions
- Reduce the number of hops and overall latency
- Distributing the load across the network
- Improving throughput of the network

The different classifications areas are presented in Table 1.

Table 1: Routing Algorithms

S.no		**Types**
S.no	Algorithm Criteria	Type
1	Based on Destinations	Unicast & Multicast[19]
2	Based on routing decision made	Centralized Routing, Distributed Routing[16] Source Routing[16], Multiphase Routing
3	Based on Adaptive Nature	Deterministic Routing , Adaptive or Oblivious Routing[17]
4	Based on Routing path adopted	Minimal & Non-Minimal routing[18]
5	Based on Network congestion	Delay & Loss

Unicast routing algorithms manages and find the path when one node desired to send the data to single node only whereas in multicast routing single node may want to send the data to multiple destinations.

In centralized routing mechanism, routing decisions are made by some dedicated centralized controller. In source Routing source node evaluates the path between source IP and destination IP and responsible for taking overall routing decisions, whereas in distributed routing the routing decisions are made by n number of nodes as packets/flits of the data moves from source to destination. Multiphase routing utilizes both source and distributed routing [24].

In Deterministic Routing strategy, the traversal path between two nodes is fixed and decided by only participating source and destination nodes whereas in adaptive or in oblivious routing , routing decisions are made by intermediate nodes and also take consideration of Network state while making decisions for the path[23,24].

According to path length criteria, the Routing algorithm may be minimal when the path determined is always shortest while traversing the packet from source to destination otherwise it is Non-Minimal [10,11].

Delay/Loss mode type algorithm are used when there is congestion in network, In delay mode data packets are never dropped, it can be delayed. Thus Guaranteed service is sure, the packet will surely reach destination. In the loss model, data packets are dropped whenever there is congestion [11, 12].

3. Flow control: Flow control determines how packets moves router to router or inter switch. Buffers are the temporary storage to store the flits while packet is in transition. Flow control techniques like Store-and-forward[6], Virtual cut-through[6] & Wormhole [6] determines the resources like channel bandwidth and buffer capacity which needs to be allocated to packet so as to achieve a optimal utilization , high throughput and good network performance.

4. Router microarchitecture: In the NoC interconnect, the communication among IP cores take places in the form of packets with the help of intelligent routers. NoC router have input ports, output ports, a crossbar switch which connects the input ports to the output ports, buffers, Processing Element (PE) corresponding to this router which all together are core of the router and performs the essential task of routing & arbitration logic. Router directs the data packets towards destination node with minimum latency.
The Design of the router is very important for overall NoC performance. It should be simple and able to perform efficiently with the power and area constraints.

Figure 1 presents the basic Mesh based NoC architecture consisting of Network Interface (NI) , Processing Element (PE), router. Mesh layout is widely accepted for NoC architectures due to its simple design and scalability characteristic.

Whenever a packet needs to be sent from a source IP to a destination IP, the packet is directed hop by hop according header information on the network based on the routing algorithm [1].

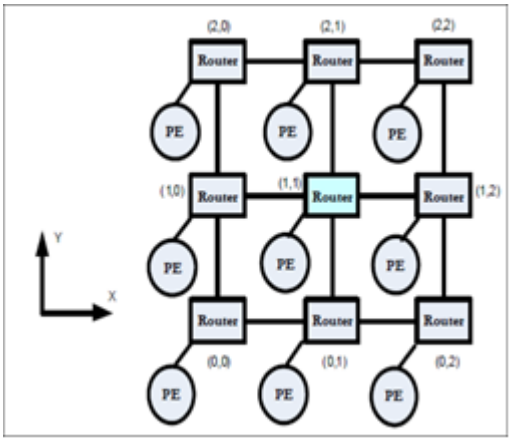

Fig 1: Basic NoC Architecture [1]

5**. Quality of Service:**
There are certain performance requirements that every NoC implementationhas tomeet. Two major categories of Quality of Services in NoC has been observed i.e Best Effort and guaranteed service [5]. Performance metrics include minimum latency, guaranteed throughput, bandwidth, path diversity and low power consumption, minimum hop Counts.

Packet Injection Rate: This is the measurable unit at which the data packets are feed into the network by a IP core known as *packet injection rate* (*pir*). Packet Injection rate varies in range of 0 and 1.For example, the *pir* of 0.2 means that each node will transmit 2 packets every 10 clock cycles.

Throughput: It is benchmark in Networks as how much data can be successfully delivered from one node to another in a specified amount of time. The unit of throughput is in bits per Second

Latency: The packet latency is also referred as end to end delay which is the total time calculated when the header flit is injected in the network by the source node till the tail flit is accepted by the destination node. It is not measured for single packet, generally average Network Latency is one of the Quality of service metrics to evaluate network performance. The packet latency consists of the transferring or admission latency, Arbitration and Routing Latency and contention Latency. Transferring or Admission latency is the time computed in forwarding the message into the network. Arbitration and Routing Latency refers to the time taken for calculating routes and for arbitration logic to select the output port and finally contention latency i.e. delay that occurred because of the contention in the routes.

Hop counts: Hop count is the number of Ip cores/routers visited by the data packet from source node to destination node[19,20]. It is the total distance traversed between sender and destination.

Drop Probability: It is the property of NoC architecture wherein data packets are dropped when there is congestion or communication load.

Path diversity: There may be multiple paths between two communicating nodes which enhances flexibility and robustness in the network for choosing the best path. It is necessary attribute which is required when there is congestion and faults and provides better load balancing within the network .The topologies and routing algorithm implemented must have features to exploit the path diversity. Mesh and Torus are the best examples which have path diversity.

III. REVIEW OF SOME CLASSIC 2D NOC TOPOLOGIES

Topology is the physical organization and interconnections between the nodes, routers and Links in the network. The two important attributes of Topologies are its Degree and Diameter. The diameter (D) can be defined as the maximal number of nodes between any pair of farthest IP core's in NoC architecture. The diameter of a network should be smaller so that it takes less time to send a message from one node to the remote node. The node degree (ND) is defined as the number of hops connected to a node. The node's I/O complexity can be determined through this attribute.

The most preferred topologies for NoC architectures are ring, mesh, torus, C-Mesh, fat tree etc [13].Mesh Topology as shown in fig.2 is a regular grid-like NoC architecture consisting of m columns and n rows.The IP core along with router are situated at the intersection of edges and easily identified with the x-y coordinates. The Major strength of mesh topology is path diversity that handles fault and scalability i.e number of IP cores can be added easily. "Chip-Level Integration of Communicating Heterogeneous Elements" CLICHÉ [6] is mesh based NOC interconnect. Another popular topology is 2D torus shown in fig.3 very similar to 2D Mesh [1, 4]. The only difference is that the wrap around edges in its corner nodes where each row and column are connected. Due to this extra wrap around edge symmetry torus has lower hop count H and has greater path diversity [7, 8] than Mesh . Cmesh is topology in which multiple nodes share same router. The fat tree is regular k-ary n-order NoC topologyin which nodes are arranged in a tree structure as shown in Fig.5. The network area, power consumption and cost of implementation for Fat tree topology is much high[13].The main advantage of this topology is its scalability [13].In our work we have evaluated these four basic topologies with respect to the parameters injection rate, latency, throughput and no.of hops.

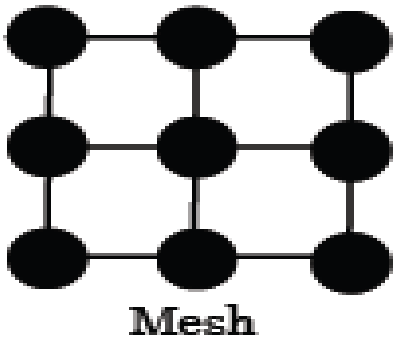

Figure 2. Mesh Topology

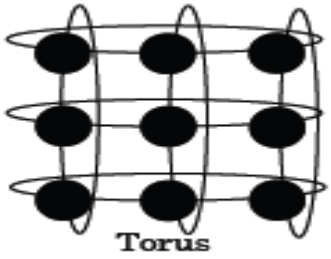

Figure 3. Torus Topology

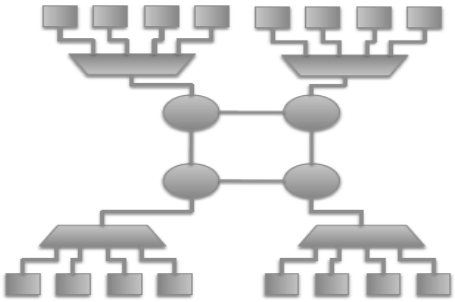

Figure 4. Cmesh (4 cores attached to 1 router)

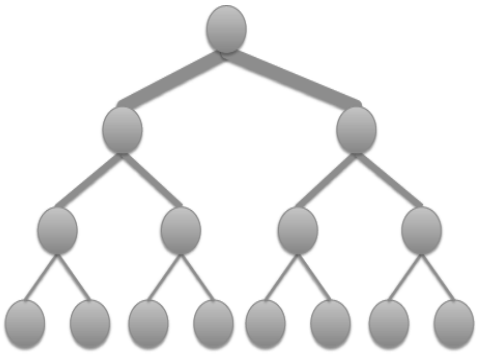

Figure 5. Fat-Tree

IV. SIMULATION AND RESULTS

In our work, we have analyze the results of simulations that have been performed using Booksim 2.0. Booksim 2.0 simulator is a detailed cycle-accurate simulator for NoC environment. It is written in C++. It can be used to model all the prime aspects including topologies, routing, and flow control, router architecture and determine the performance parameters[15].

We have analyzed the impact of injection rate on the parameters Latency and Throughput for the 4 basic topologies i.e Torus, Mesh , fat-tree & c-mesh. The various simulation parameters used are shown in table 2 below:

Table 2: Parameters used for simulation

Number of Nodes	16
Routing Algorithm	DOR Routing (XY)
Traffic Pattern	Uniform
Virtual Channels	8
Number of Buffers	8
Simulation type	Latency and throughput against injection rate

With varying injection rate the latency and throughput are simulated for all four topology. The results obtained are the average values obtained after run of 20 cycles of booksim 2.0 which are shown in the following table and figures.

Table 3. Injection Rate vs Latency

Injection rate	Latency (Mesh)	Latency (torus)	Latency (cmesh)	Latency (fat tree)
0.1	8.16688	6.09203	11.1967	23.22
0.2	8.36785	6.12798	11.3399	23.2502
0.25	8.57567	6.22626	11.5578	23.4255
0.3	8.68573	6.29864	11.921	23.5276
0.35	8.8553	6.36982	12.612	23.7092
0.4	9.12075	6.48916	14.2959	23.8873
0.45	9.39171	6.58926	35.6894	24.0946
0.5	9.72269	6.75767		24.3976
0.55	10.1946	7.00264		24.7219
0.6	10.8985	7.26118		25.2352
0.65	11.6745	7.66861		25.8901
0.7	13.0366	8.21282		26.9072
0.75	15.243	9.11454		28.2196
0.8	18.7931	10.6792		31.537
0.85		14.5583		39.1369

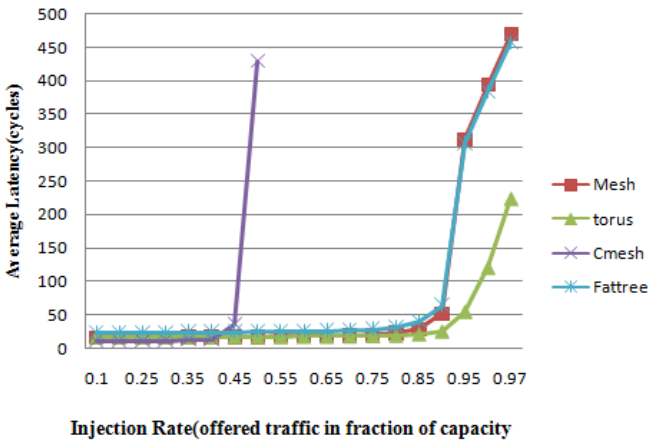

Figure 6. Network Latency comparison

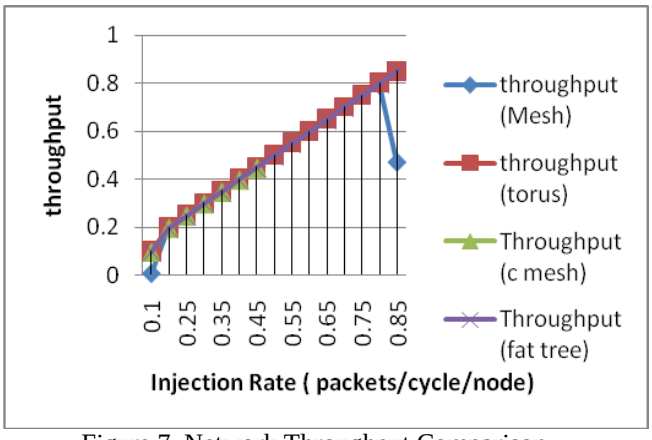

Figure 7. Network Throughput Comparison

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 4. Injection Rate vs Throughput

Injection rate	throughput (Mesh)	throughput (torus)	Throughput (c mesh)	Throughput (fat tree)
0.1	0.009941	0.101265	0.0993333	0.0994167
0.2	0.201435	0.201753	0.199437	0.199687
0.25	0.252303	0.252425	0.25	0.25
0.3	0.302149	0.302415	0.299667	0.3
0.35	0.352145	0.353373	0.348479	0.348971
0.4	0.402233	0.40228	0.398188	0.398417
0.45	0.452269	0.452514	0.447271	0.45
0.5	0.501804	0.502299		0.499063
0.55	0.552312	0.552439		0.548708
0.6	0.601104	0.601128		0.598896
0.65	0.650265	0.651037		0.64875
0.7	0.699069	0.700533		0.699042
0.75	0.749401	0.750474		0.749708
0.8	0.79271	0.800443		0.799687
0.85	0.47117	0.849584		0.851771

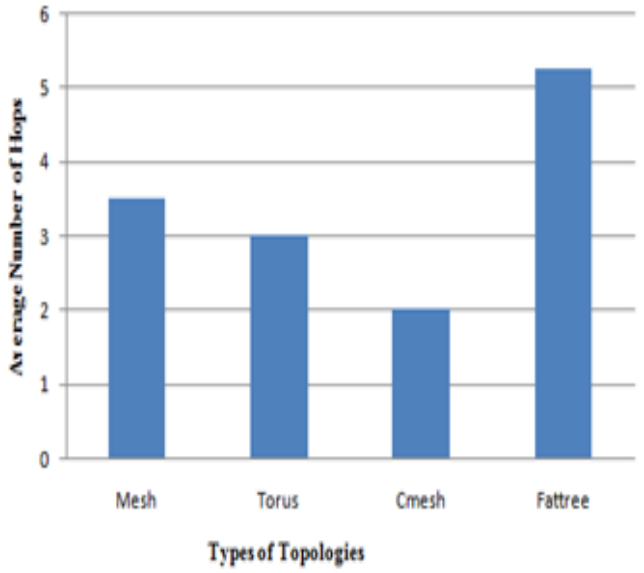

Figure 8 Hops comparison

V. CONCLUSION

Networks-on-Chip are leading edge technology due to growing requirement in embedded systems and emerged as an efficient and scalable communication infrastructure able to accommodate large numbers of IP cores on single chip. We have presented the performance of four basic topologies Mesh, Torus, C-Mesh & Fat-tree. The results are presented in various graphs obtained from simulation and it can be visualized from figure 6 and table 3 the latency increases with the increasing values of injection rate. Torus performs better among all the evaluated topologies in the terms of latency and it is minimum followed by mesh topology. Torus has minimum latency because of wrap around edges but wire length is increased and hence the power consumption also increases. Fat-tree topology shows highest latency among all 4 topologies.From Figure 7 and table 4, it can be concluded from the experiments that as the throughput increases the injection rate also increases linearly. Injection rate range varies for every topology even for a same size network. From this experimental results and values we can say Cmesh topology do not support high injection rate range. Average number of Hops comparison is shown in fig 8. It is clearly visible that Torus has better performance in comparison with all other topology for each performance parameters and can be suitable choice for future NOC architectures.

This Work can further be extended to perform the analysis of various topologies with varying simulation parameters like routing algorithm and with various traffic pattern.We want to further extend this work to explore 3D NoC architectures performance. 3D architectures are susceptible to congestion due to large number of IP cores , thus the fault-tolerance and reliability of these networks are very critical issue for future applications.

REFERENCES

1. Dally, William J., and Brian Towles. "Route packets, not wires: on-chip interconnection networks." Design Automation Conference, 2001. Proceedings. IEEE, 2001.
2. Booksim2.0https://nocs.stanford.edu/cgibin/trac.cgi/wiki/Resources/Booksim
3. W. J. Dally and B. Towles, “Principles and Practices of Interconnection Networks. San Francisco”, CA: Morgan Kaufmann, 2004.
4. Nan Jiang, Daniel U. Becker, George Michelogiannakis, James Balfour, Brian Towles, John Kim and William J. Dally, “A Detailed and Flexible Cycle-Accurate Network-on-Chip Simulator”. In the Proceedings of the 2013 IEEE International Symposium on Performance Analysis of Systems and Software, pp. 86-96, 2013.
5. B. Gebremichael, F.W. Vaandrager and M. Zhang, “Formal Models of Guaranteed and Best-Effort Services for Networks on Chip”, Technical Report ICIS-R05016, ICIS, Radboud University Nijmegen, March 2005
6. Kumar, S., Jantsch, A., Soininen, J. P., Forsell, M., Millberg, M., Oberg, J.&Hemani, A. (2002). A network on chip architecture and design methodology. In VLSI, 2002. Proceedings. IEEE Computer Society Annual Symposium on (pp. 105-112). IEEE.
7. Bjerregaard, T., & Mahadevan, S. (2006). A survey of research and practices of network-on-chip. ACM Computing Surveys (CSUR), 38(1), 1.
8. Mirza-Aghatabar, M., Koohi, S., Hessabi, S., &Pedram, M. (2007, August). An Empirical Investigation of Mesh and Torus NoC Topologies Under Different Routing Algorithms and Traffic Models. In DSD (Vol. 7, pp. 19-26).
9. J. Liang, S. Swaminathan, and R. Tessier, "aSOC: AScalable, Single-Chip communications Architecture", InIEEE International Conference on Parallel Architectures andCompilation Techniques, Oct. 2000, pp. 37-46.
10. L. M. Ni and P. K. McKinley, "A Survey of Wormhole Routing Techniques in Direct Networks", IEEE Computer Magazine, v.26(2), February, 1993, pp. 62-76.
11. M.A. Kinsy, M.H. Cho, T. Wen, G.E. Suh, M.V. Dijk, and S. Devadas, "Application-aware deadlock-free oblivious routing", in Proc. ISCA, 2009, pp.208-219.
12. H. Zhu, P. P. Pande, and C. Grecu, "Performance evaluation of adaptive routing algorithms for achieving fault tolerance in NoC fabrics", in IEEE International Conf. on Application-specific Systems, Architectures and Processors, ASAP 2007, Montreal, Que., Jul. 2007, pp. 42–47.
13. P. Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh, Performance evaluation and design trade-o_s for network-on-chip interconnect architectures," Computers, IEEE Transactions on, vol. 54, no. 8, pp. 1025{1040, Aug 2005.

14. Jiang, N., Balfour, J., Becker, D. U., Towles, B., Dally, W. J., Michelogiannakis, G., & Kim, J. (2013, April). A detailed and flexible cycle-accurate network-on-chip simulator. In Performance Analysis of Systems and Software (ISPASS), 2013 IEEE International Symposium on (pp. 86-96). IEEE.
15. Booksim 2.0, http://nocs.stanford.edu/booksim.html, 2012.
16. Achballah, A. B., & Saoud, S. B. (2013). A survey of network-on-chip tools. arXiv preprint arXiv:1312.2976.
17. Samman, F. A., Hollstein, T., & Glesner, M. (2009). Networks-on-chip based on dynamic wormhole packet identity mapping management. VLSI Design, 2009, 2
18. Rantala, V., T. Lehtonen, J. Plosila, 2006. “Network on Chip Routing Algorithms”, TUCS Technical Report, pp. 779.
19. Ville Rantala, Teijo Lehtonen and Juha Plosila, "Network on Chip Routing Algorithms", University of Turku, Department of Information Technology Joukahaisenkatu 3-5 B, 20520 Turku, Finland, TUCS Technical Report, August 2006.