International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 5, No.2 (March-2016)
http://dx.doi.org/10.12785/ijcds/050203

# Performance Evaluation of Modern Network-on-Chip Router Architectures

**Jawwad Latif[1], Sadia Azam[2], Hassan Nazeer Chaudhry[3] and Tahir Muhammad[4]**

[1,2,4] *Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan*
[2] *Department, Technische Universität Darmstadt, Darmstadt, Germany*

*Received 20 July 2015, Revised 14 December 2015, Accepted 17 January 2016, Published 1March 2016*

**Abstract:** On chip interconnection networks simplify the challenges of integrating large number of processing elements. Routers are backbone of networks. Buffers and crossbar in router consumes significant area and power of network. They have huge impact on overall performance and cost of network. Dual Xbar router architecture combines buffered and bufferless feature to reduce buffer read/write energy with dual crossbars. While Switch folding technique introduced to reduce wire density and decrease muxes in crossbar by increasing resource utilization. In this paper, we propose Folded Dual Xbar architecture by combining the Dual Xbar and Folding technique in order to get advantages of both architectures. Performance of architectures is implemented and evaluated using OMNET++ platform by applying multiple traffic patterns under different load conditions. We further calculated buffered and bufferless events to estimate the reduction in buffer read/write energy. Simulation results shows that there is slight increase in throughput and reduction in buffer read/write energy by average 46% at high loads in proposed 2-Folded Dual Xbar as compared to conventional architecture. Proposed 3-Folded Dual Xbar results at least 16.6 % increase in throughput as compared to conventional architecture with 43-45% reduced buffer read/write energy but slight increase in crossbar. Throughput of 3-Folded Dual Xbar decreased only by 5-7% as compared to Dual Xbar with distributed wire density advantage..

**Keywords:** Network-on-Chip, Dual Xbar; Folding technique; Router architecture

## 1. INTRODUCTION

Due to the advancement in deep-submicron technology, processing elements on chip are increasing with the passage of time [1]. Traditional System-on-Chip (SoC) technology uses dedicated wires for communication between IP cores. The trend of increasing the processing elements on chip raises communication complexity and affects the scalability of on chip systems. The parasitic effects becomes dominant with large number of interconnects [2]. Cost of SoC system is highly influenced by interconnects which pays important role in power consumption, performance and overall size of the system. To minimize the limitations of SoC technology, Network-on-Chip (NoC) is introduced to improve communication infrastructure in SoC [3]. In NoC, the dedicated wires for interconnection are being replaced by network with few numbers of links and routers to route the packets from source to destination. NoC offers flexibility to integrate large number of IP cores on chip [5].

Network-on-chip technology offers an attractive solution to keep up with the increasing performance requirements. Efficient design of NoC systems makes use of scarce communication resources to provide high bandwidth, low latency, reduced power consumption and minimum area requirement. One of the main features of NoC approach is that it provides high degree of Scalability. In traditional SoC approach increasing number of number of transistor on chip doesn't offer much scalability while NoC approach offers an easy way to integrate more number of system components on a chip. In complex engineering systems, modular approach is used to divide the design problem in different smaller problems and with well-defined interface. With NoC technology approach the design problem is instead of "How to design flexible interconnect for a complex system? But its "How to design best routing and transportation of packets?". Decoupling of layers has optimized the implementation of each layer independently. For example, designer can implement different topologies at transport layer without changing the rest of layers [5]. Globally Asynchronous Locally Synchronous (GALS) architectures are made up of locally synchronous modules which are connected through globally asynchronous network. This concept breaks the single clock domain into multiple clock domains.

*E-mail: jaidpk@hotmail.com, sadia.azam@hitecuni.edu.pk, hassan.bhatti@uettaxila.edu.pk, tm.uettaxila@yahoo.com*

Distributed nature of NoC is well suited by multiple clock domains. GALS concept increases the flexibility and also eliminate the issues of clock distribution in large chips. We can apply this concept in NoC in easy way to get its benefits [6].

Main components of NoC architecture are links, network Interface and router. Links are used to physically connect the nodes and move the data between nodes, Network Interface provides interface between network and IP cores while Router implements the communication protocol. Routers are the backbone of networks and their implementation cost (area/power) determines the cost of whole network. Crossbar and buffers are the core components of router which consumes significant portion of the router's area and power consumption. Different architectures of router have been proposed to improve the performance of the network, reduce the power consumed by buffers and crossbar, and area occupied by them.

Recently, Dual Xbar and Switch folding technique were introduced for router architecture. In Dual Xbar significant amount of read/write buffer power has been reduced by combining buffered and bufferless in router architecture with 20% increase in throughput but utilizes two crossbars which increase area of router [7]. In Switch folding technique, time multiplexing concept is implemented to reduce wire density and the area of router by increasing resource utilization [8]. In order to get the benefits of both techniques, we propose Folded Dual Xbar technique to get the main advantages of both architectures. Folded Dual Xbar contains buffered and bufferless feature to save read/write energy consumption and folding technique to distribute wire density and get area benefits as well. The major contributions in this paper are performance evaluation and comparison of Conventional architecture, Dual Xbar and Folded Dual Xbar by applying Uniform Random and Hotspot traffic patterns under different load conditions. Buffered and bufferless events are also calculated to estimate the reduction in buffer read/write energy.

## 2. RELATED WORK

Buffers and crossbar are the most important components in routers architecture. They consume significant portion of area and power in a network which determines the cost of the system. Many techniques have been proposed to increase the performance while keeping power consumption of buffer minimum. Buffers in router architecture can be at output and input or at both sides. The Output-buffered router (OBR) gives high saturation throughput than Input-buffered router (IBR) under high offered load. Distributed-shared buffer (DSB) [9] was proposed in which author emulated OBR based on distributed-shared buffer architecture of router. This mechanism uses two crossbars and memory elements are placed between them. . For synthetic traffic pattern, DSB router [10] has 19% increase in saturation throughput and more than 94% ideal saturation throughput. Increase in throughput gives low network latency. For SPLASH-2 benchmark, DSB router has achieved 60% reduction in network latency. Mechanism of bypassing the router to reduce per hope delay was introduced earlier using traditional router. This technique although reduces per hop latency but increases the complexity and cost of router. Low-Cost Router Micro architecture [11] reduces the complexity and the cost of router while maintaining the bypass mechanism to reduce per hope latency.

Enhanced design space of two stage Elastic buffers (EB) [12] was proposed in which cycle time has been reduced by 42% as a result maximum throughput is achieved. The area of two-stage router is reduced by replacing three-slot output EB with two-stage EB and parallel processing of Look-ahead routing mechanism. Reduction in area is almost 20 % as compared to three-slot output EB router. Many techniques have been proposed for efficient utilization of buffers to reduce the power consumption and area consumption. ViChar (A Dynamic Virtual Channel Regulator for Network-on-Chip Routers) is one of them in which buffer utilization is made efficient by dynamically allocation buffer resources according to the condition of traffic pattern in the network [13]. Bufferless routing algorithm eliminates buffers which results in significant reduction in overall power consumption of network [14]. Performance in bufferless on-chip networks degrades due to increase in deflection of packets and contention at high loads. Deflection of packets also reduces the power benefits in bufferless router networks. Clumsy Flow Control (CFC) [19] can reduce amount of routing deflection up to 92% at high loads and also reduces power consumption by approximately 39% as compare to baseline router. Configuring algorithm [20] is designed for application specific NoC which is based on reconfigurable on-chip architecture. It increases flexibility in topology generation-floorplanning scheme and mapping scheme, and also reduces design complexity. Efficient utilization of resources improves the performance of system. Designers confront major problems in designing optimal application mapping and scheduling. Models of scheduling and applications mapping are developed using Constraint Programming (CP) [21]. Analysis indicates that the optimization problems of scheduling and application mapping can be solved under shorter run-time limits as compared to Mixed Integer Programming (MIP) models.

An iDEAL architecture [15] uses dual function (three-state repeaters) links, which can be used for transmission and data storage. When there is congestion in the network, three state repeaters hold the data until congestion alleviated. With adaptive dual function links approximately 30% savings in overall network power and 40% savings in buffer power have been achieved with only 1-3% drop in performance. SCARAB router

architecture [16] was introduced to maintain the power and advantages of bufferless router with efficient results. Different mechanisms were proposed to support dropping protocol in order to reduce packet drops and retransmission of packets. Analysis of Blind Packet Switched (BPS), SCARAB router and Hot-Potato (HP) router is done in this paper. SCARAB has shown 12.6% more efficiency than BPS and 18.3% more than HP. The main drawback of SCARAB is the area overhead. During execution time not all routers have packets to store. Buffer utilization can be made more efficient by sharing buffers between ports. In RoShaq architecture, [17] buffer utilization was improved by sharing multiple buffer queues between input ports. This technique results in 14 % increase in saturation throughput over traditional router. The idea in RoShaq is to share buffers which resulted in higher throughput at high traffic loads. Performance, power and energy evaluation of multiple architectures are also discussed [22]. QoS-aware and Congestion Aware router architecture [23] provides quality-oriented transmission and improves throughput by balancing traffic load over network. Performance is improved for advanced mesh NoC platforms with negligible cost overhead.

Reducing the size of buffers result in performance degradation, while increasing the buffers size result in high power and area consumption. DXbar [7]: an innovative dual-crossbar design has been proposed by combining the advantages of buffered and bufferless router. DXbar is implemented using two crossbars. One crossbar has bufferless inputs which named as primary crossbar and other crossbar has buffered inputs named as secondary crossbar. There is at least 20% performance improvement in terms of throughput and latency, and at least 20% power saving over buffered networks with virtual channels [7]. The use of two crossbars causes the most area overhead in DXbar. Time division multiplexing technique [8] is adopted in crossbar to reduce the area, cost and wire density in a router. In this mechanism the output ports are folded to implement portion of router's arbiters and multiplexers.

## 3. IMPLEMENTATION

### A. *Network Model*

Network of all three architectures consists of cores and routers. Routers are connected thorough 2D-Mesh topology. Each router is also connected with a core as shown in Fig. 1. Core consists of source and sink. Source module generates packets depending on clock and adds necessary information to the packet. When the packet reaches at its destination, it is forwarded to the sink module where some statistics are recorded. Router implements the routing of packets by implementing XY routing, VC calculation, arbitration, credit sharing etc... Three architectures of router are implemented which are elaborated in next sections. For analysis, Uniform Random and Hotspot traffic patterns are applied on 4x4 mesh size network with the channel bandwidth of 16Gbs.

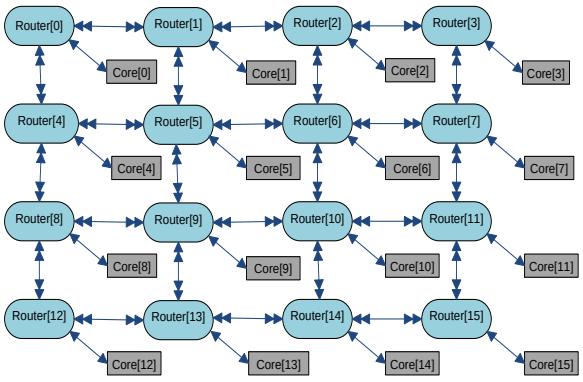


Figure 1. Network Model

*1) Core Module:* Applications are executed on IP cores. They generate packets and receive packets from other IP cores. In our network, Core is implemented to exchange packets. It contains two sub-modules Source and Sink.

*a) Source: It* generates packets according to arrival time (clock). Offered load depends on the value of arrival time. For analysis, different values of arrival time have been taken and average throughput is measured for all three architectures. Source adds necessary information to the packet including source address and destination address. Packet formats in conventional, Dual Xbar and Folded Dual Xbar router networks are given in Fig. 2. Packet format in source module of conventional routers network, wormhole routing is implemented in which packet is divided into flits. There are three types of flits start flit, mid flit and end flit. Flit type field is added in the packet format to differentiate between the types of flits. While in Dual Xbar and Folded Dual Xbar, there is no division of packets into flits. Each flit is considered as a complete packet as shown in Fig. 2. The reason is both architectures cannot support in order delivery of flits due to the nature of their architecture.

*b) Sink:* When packet reaches at its destination it is forwarded to sink module. Sink module records some statistics like end2endlatancy of packets, number of packet received etc...

### B. *Proposed Architecture*

In this paper, proposed architecture is Folded Dual Xbar. The idea is to combine two techniques Dual Xbar and Switch folding technique. Dual Xbar has combined buffer and bufferless feature while Switch folding gives area benefits and reduce wire density by time division multiplexing and resource utilization. Proposed architecture Folded Dual Xbar utilizes both concepts. It is being implemented on OMNeT++ Platform.


Figure 2(a). Packet format in Conventional router


Figure 2(b). Packet format in Dual Xbar router


Figure 2(c). Packet format in Folded Dual Xbar

It has buffer and bufferless feature like Dual Xbar and time division multiplexing of Switch folding technique. Therefore, this architecture gets the benefits of reducing buffer read/write energy while having benefits of Switch folding mechanism. Design of 2-Folded Dual Xbar is shown in Fig. 3(a). It has dual folded crossbars. One is primary folded crossbar which has bufferless connections with input ports. When packet arrives and corresponding output port is idle then router allows it to pass through primary folded crossbar which has bufferless feature. Second is secondary crossbar which has buffers connected to its input ports. When packet arrives and its corresponding out port is busy then the packet is stored in buffer and passes through secondary crossbar after receiving permission from scheduler. Crossbars are also folded in which it increases resource utilization through time division multiplexing. When there are multiple packets looking for same output port it follows time division multiplexing. Packets are not divided into flits because the architecture cannot follow in order reception of packets. Every packet has a unique packet ID. Implementation of 2-Folded Dual Xbar is described in flowchart shown in Fig. 3(b). Router receives incoming packet and forward it to output port calculation module. XY routing is used to determine the out port.

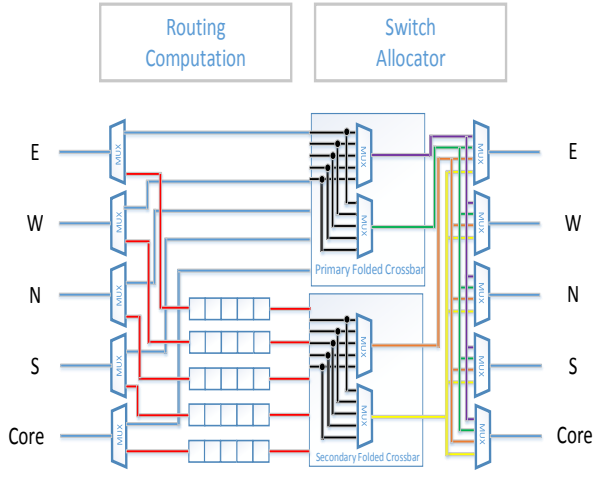


Fig. 3(a). 2-Folded Dual Xbar architecture

After receiving the out port value from XY routing module, packet checks whether the corresponding out port is busy or free. A single cycle plenty can be reduced by using look ahead routing for output port calculation as implemented. If the corresponding output port is idle than packet is moved through primary folded crossbar. If the output port is busy than packet is stored in FIFO input buffers and request is forwarded to the arbitration unit. Arbiter implements round robin scheduling algorithm. When the out port becomes free, scheduler sends grant to inport unit. Inport unit receives grant and moves the corresponding packet through secondary folded crossbar. If any situation arises in which primary folded crossbar continuously send packets, causing packets in buffer to wait for long duration then fairness between secondary packets and primary packets can also be applied by using appropriate value of counter. In network, crossbar failure is possible. Fault Tolerance at hardware level can be applied in Dual crossbar architecture6. If any crossbar fails, the routing could be shifted to either crossbar.

## 4. EVALUATION

Our analysis includes the performance evaluation of three router architectures. 1) Conventional architecture 2) Dual Xbar and 3) Folded Dual Xbar. HNOCS [16] simulator is used for conventional router. Folded Dual Xbar further consists of different number of folding switches. 2-Folded Dual Xbar contains two muxes in each crossbar and 3Folded contains 3 muxes in each crossbar. Comparison of 2-Folded Dual Xbar. Dual Xbar, 3-Folded Dual Xbar and conventional router is being done. Throughput and latency graphs are plotted by applying Hotspot traffic pattern and Uniform Random traffic pattern. In order to estimate buffer read/power reduction, buffered and bufferless events are also counted and shown in Table. 1. Analysis is done on 4x4 2D mesh topology network using OMNeT++ platform.

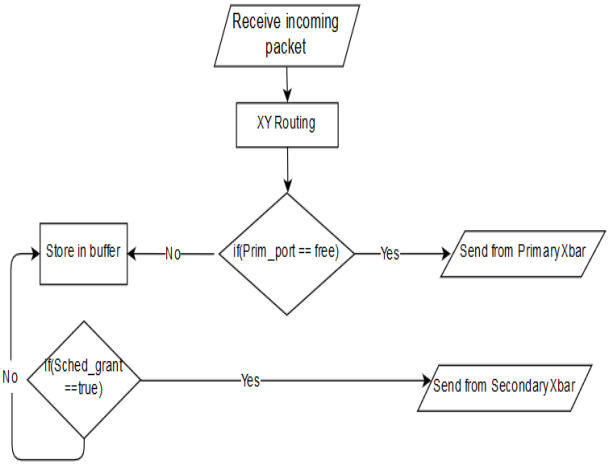


Fig. 3(b). Packet routing in Folded Dual Xbar

*A. Uniform Random Traffic Pattern*

We compare conventional architecture with 2-Folded Dual Xbar. In our analysis, conventional router contains 2 virtual channels with a capacity to store 5 flits. Each packet is divided into 10 flits. The size of each flit is 4Bytes. On the other hand, there is no division of packets in 2-Folded Dual Xbar network. Each flit is considered as a packet and each packet is of 4Bytes. 2-Folded Dual Xbar architecture consists of one FIFO buffer with a capacity of storing 10 flits. Uniform Random traffic pattern is applied and value of throughput and latency is measured on multiple offered loads. To estimate that how much buffer read/write events are reduced in 2-Folded Dual Xbar, buffered and bufferless events are also counted and shown in Table. 1. The graph showing throughput comparison between conventional architecture and 2-Folded Dual Xbar is shown in Fig. 4. Throughput is calculated from equation 1.

$$Throughput = \frac{\sum_1^N R.P * P.L * 4}{T.C * S.T * 10^9} [GB/s] \quad (1)$$

N = Total number of cores
R.P = Received packets
P.L = Packet length in flits
T.C = Total number of cores
S.T = Total simulation time

Offered load is applied from 0.5 to 1.3 where both networks get saturated. They have shown almost same behaviour on all offered loads except a slight increase in throughput by 2-Folded Dual Xbar at saturation. Conventional router with 2 virtual channels has saturated throughput 1.02 and 2-Folded Dual Xbar network get saturated at 1.07. The main advantage of 2-Folded Dual Xbar over conventional router is the bufferless feature to reduced buffer read/write energy. The crossbar in 2-Folded contains total 4 switch elements with slight overhead of distributor circuit at outputs of crossbar. While conventional router has 5 switch elements in its crossbar. All flits in conventional are always stored in buffer/vc while 2-Folded Dual Xbar reduces buffer read/write energy by directly sending the packet through primary crossbar when the corresponding output is idle. This feature is adopted from Dual Xbar6. Switch folding7 is applied to reduce wire density and muxes by increasing resource utilization. The percentage of bufferless events at high loads 1.06, 1.14 and 1.33 are 51%, 45.6 % and 40%. An average of 46% buffer read/write energy is saved at high loads in 2-Folded Dual Xbar as compared to conventional architecture with almost same throughput. On other hand, the wire density between the input ports and crossbar also decreased in 2-Folded Dual Xbar as compare to conventional architecture which improves layout density of the circuit [7]. Latency of every packet is measured and then converted into average latency per byte through the equation given as:

$$Latency = \frac{\sum_1^N L.R.P}{\sum_1^N B.R} [nsec/byte] \quad (2)$$

N = Total number of cores
L.R.P = Latency of received packets
B.P = Number of byte received

Latency graph is also plotted as shown in Fig. 5. Graph shows the behaviour of latency in conventional and 2-Folded Dual Xbar architecture. Offered load is applied from 0.5 to 1.3 where both networks get saturated. The values of throughput are same from 0.5 to 0.94 offered load so the latency values as well. At 0.94 offered load, there is a slight increase in latency of 2-Folded Dual Xbar because its throughput is less by 0.01 from conventional router. Interesting to see that only at 1.0 offered load the latency of 2-Folded Dual Xbar is more than conventional router which could be the result of worst case or average case scenario in 2-Folded Dual Xbar in which all four input ports of router receives packet to forward them in all four different directions. After 1.0 offered load the latency of 2-Folded Dual Xbar decreases because its saturated throughput is 1.07 while conventional router has almost 1.05 throughput value. Throughput and Latency comparison of three architectures conventional, 3-Folded Dual Xbar and Dual Xbar is also shown in Fig. 6. The specifications are same as described in previous comparison. 3-Folded Dual Xbar contains 6 switches in its crossbar with slight overhead of distribution circuit and conventional router contains 5 switches in its crossbar. According to measured statistics in OMNET++, 3-Folded Dual Xbar has 16.6% increases in throughput as compared to conventional router having 2 virtual channels. Dual Xbar and 3-Folded Dual Xbar has bufferless feature while conventional router always stores flit into buffer. At saturation level, 43 – 45% events are bufferless in 3-Folded Dual Xbar which reduces significant amount of buffer read/write energy. We have achieved 16.6% increase in throughput and 43-45% saves in buffer read/write energy as compared to conventional router with 2 virtual channels. Dual Xbar has saturated throughput 1.37 which is 23.3% more than conventional router and 8% more than 3-Folded Dual Xbar. Dual Xbar also reduces buffer read/write energy by 52% at saturation level. Dual Xbar has two crossbars which consumes significant area and increases wire density between crossbar and input ports. So, we have more options for area selection through different folding technique with buffer power reduction as well. Latency graph is also shown in Fig. 7. It shows similar trend as expected from throughput graph. Traditional router's latency with 2 virtual channels continues to increase after 0.94 offered load. Increase in throughput of Dual Xbar resulted in decrease in latency as compared to 3-Folded Dual Xbar after 1.14 offered load.

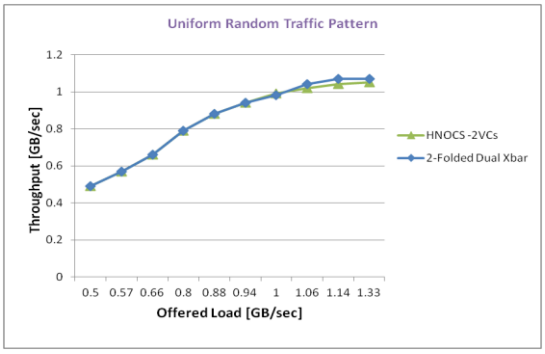


Fig. 4: Throughput of Uniform Random traffic pattern

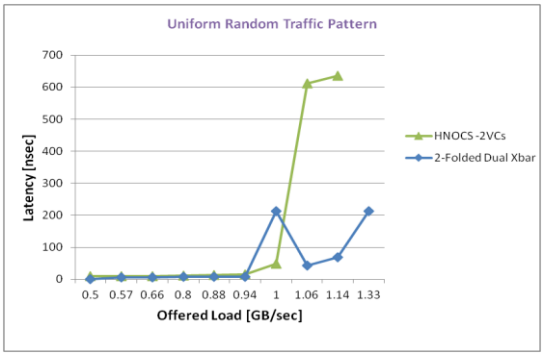


Fig. 5: Latency of Uniform Random traffic pattern

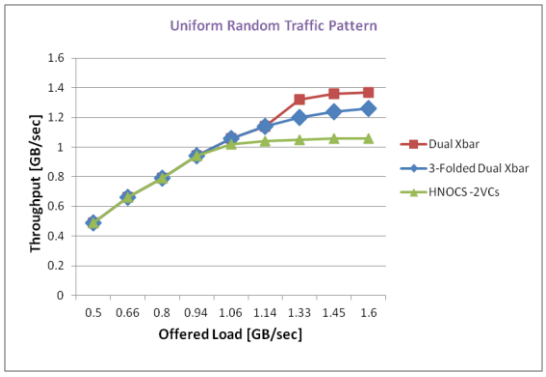


Fig. 6: Throughput of Uniform Random traffic pattern

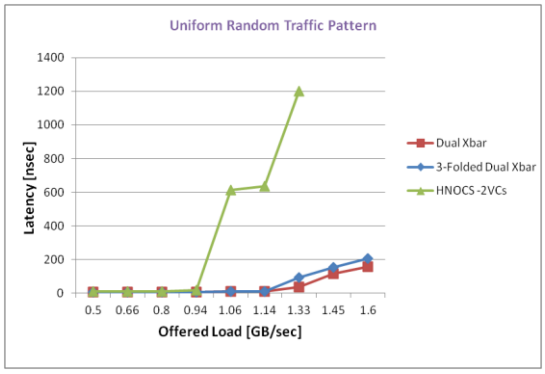


Fig. 7: Latency of Uniform Random traffic pattern

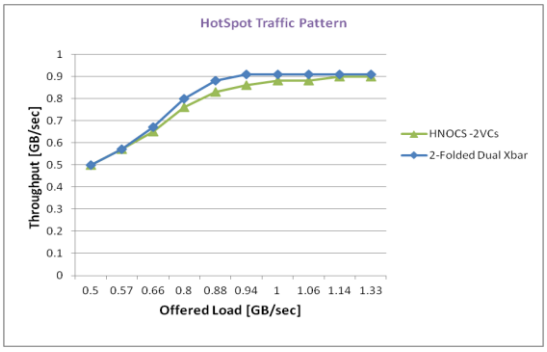


Fig. 8: Throughput of Hotspot traffic pattern

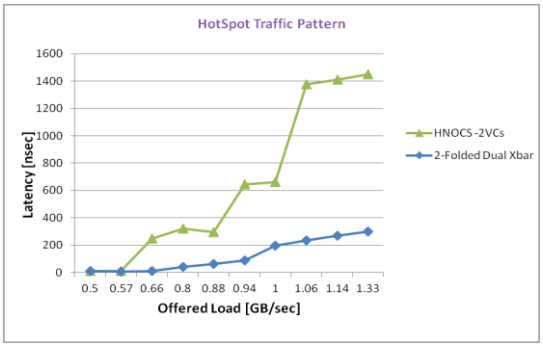


Fig. 9: Latency of Hotspot traffic pattern

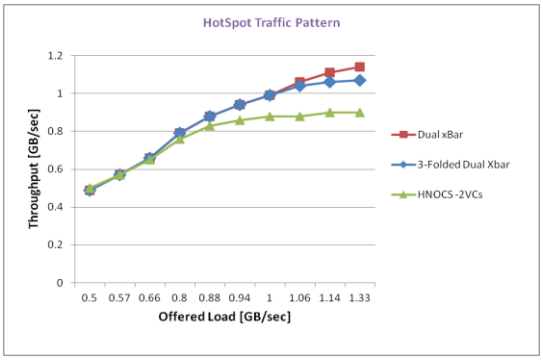


Fig. 10: Throughput of Hotspot traffic pattern

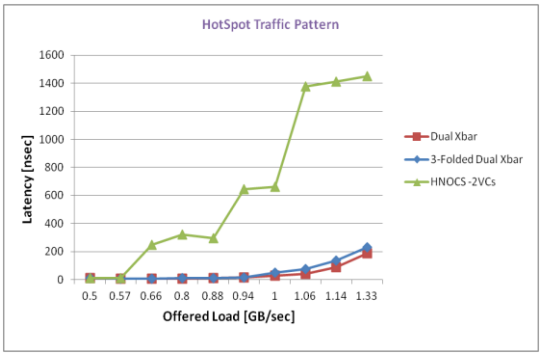


Fig. 11: Latency of Hotspot traffic pattern

### B. *Hotspot Traffic Pattern*

Specifications of packets size and virtual channels are same as in Uniform Random. The throughput comparison between conventional and 2-Folded Dual Xbar is shown in Fig. 8. Offered load is applied from 0.5 to 1.3 where both networks get saturated. The final saturated throughputs of both networks are same but the interesting point is that the 2-Folded Dual Xbar reaches its saturation level too early than conventional router. At 0.94 offered load, 2-Folded Dual Xbar network reaches at its saturation level having throughput 0.91 and remains same on all other higher offered loads. It gives greater throughput than conventional router with 2 virtual channels on almost all offered loads except at saturation level where they both have same value of throughput. At saturation level, 2-Folded Dual Xbar has ability to save 40% buffer read/write energy. Latency graph is also shown in Fig. 9. Conventional router has shown significant increase in latency. Throughput performance of three architectures is shown in Fig. 10. 3-Folded Dual Xbar has 15.8% increases in throughput as compared to conventional router. At saturation level, 47 – 50% events are bufferless in 3Folded Dual Xbar which saves significant amount of buffer read/write energy. We have achieved 15.8% increase in throughput and 47-50% saves in buffer read/write energy as compared to conventional router. Dual Xbar has saturated throughput 1.14 which is 21% more than conventional router and 6% more than 3-Folded Dual Xbar. Dual Xbar reduces buffer read/write energy by 55% at saturation level Latency graph is also shown in Fig. 11. Conventional router's latency continues to increase after 0.57 offered load. Increase in throughput of Dual Xbar resulted in decrease in latency as compared to 3-Folded Dual Xbar. The saturation throughput in conventional router, 2-Folded Dual Xbar, 3-Folded Dual Xbar and Dual Xbar is decreased by 14.2%, 15%, 15% and 16.7%. Comparison of all three architectures is summarized in Table 1.

## 5. CONCLUSION AND FUTURE WORK

Our work briefly compares the performance of multiple architectures using OMNeT++ platform and also estimated possible parameters to compute energy consumption of buffers. In future, we are interested to do comprehensive analysis of area and power using ORION 3.0 [24]model. Furthermore, these architectures can also be evaluated on different topologies and routing mechanisms to find out the better performance for specific applications. Recently, different architectures are also proposed to increase performance and, reduce area and power. Comprehensive analysis and comparison between our proposed and recently introduced architectures could be done in terms of performance, size and energy.

Table 1. Performance summary

| R.Arch | Uniform Random | | | Hotspot | | |
|---|---|---|---|---|---|---|
| | Max. TP | B.E % | # X.M | Max. TP | B.E | # X.M |
| H-2VC | 1.05 | 0 | 5 | 0.9 | 0 | 5 |
| 2-FDX | 1.07 | 40 | 4 | 0.91 | 40 | 4 |
| 3-FDX | 1.26 | 44 | 6 | 1.07 | 45 | 6 |
| DX | 1.37 | 52 | 10 | 1.14 | 55 | 10 |

R.Arch = Router Architecture
Max. TP = Maximum Throughput
B.E = Bufferles Events
#X.M = No. of Crossbar Mux's
H-2VC = Conventional Architecture-2VCs
2-FDX = 2-Folded Dual Xbar
3-FDX = 3-Folded Dual Xbar
DX = Dual Xbar

## Acknowledgment

This research has been done at ASIC lab , University of Engineering and Technology Taxila as a thesis for the award of degree of MS Computer Engineering. Our big regards and thanks to University Of Engineering & Technology, Taxila for providing a remarkable research environment for Full time MSc. Students.

## REFERENCES

[1] Bouhraoua, Abdelhafid, and Muhammad ES Elrabaa. "IMPROVED MODIFIED FAT-TREE TOPOLOGY NETWORK-ON-CHIP." Journal of Circuits, Systems, and Computers 20.04 (2011): 757-780.

[2] Chan, Cheng-Hao, et al. "A priority based output arbiter for NoC router." Circuits and Systems (ISCAS), 2011 IEEE International Symposium on. IEEE, 2011.

[3] J. Handy, "NoC interconnect improves SoC economics," Objective analysis - Semiconductor market research, 2011.

[4] Onizawa, Naoya, et al. "High-Throughput Compact Delay-Insensitive Asynchronous NoC Router." Computers, IEEE Transactions on 63.3 (2014): 637-649.

[5] "Applying the Benefits of Network on a Chip Architecture to FPGA System Design" White Paper , April 2011, Altera Corporation

[6] Cheshmi, Kazem, et al. "Quota setting router architecture for quality of service in GALS NoC." Rapid System Prototyping (RSP), 2013 International Symposium on. IEEE, 2013.

[7] Zhang, Yixuan, Randy Morris Jr, and Avinash K. Kodi. "Design of a performance enhanced and power reduced dual-crossbar Network-on-Chip (NoC) architecture." Microprocessors and Microsystems 35.2 (2011): 110-118.

[8] G. Dimitrakopoulos , N. Georgiadis, C. Nicopoulos and E. Kalligeros "Switch Folding Technique: Network-on-Chip with time multiplexed output ports" DATE 2013

[9] Soteriou, Vassos, et al. "A high-throughput distributed shared-buffer NoC router." Computer Architecture Letters 8.1 (2009): 21-24.

[10] Ramanujam, Rohit Sunkam, et al. "Design of a high-throughput distributed shared-buffer NoC router." Networks-on-Chip (NOCS), 2010 Fourth ACM/IEEE International Symposium on. IEEE, 2010.

[11] Kim, John. "Low-cost router microarchitecture for on-chip networks."Proceedings of the 42nd annual IEEE/ACM international symposium on microarchitecture. ACM, 2009.

[12] Michelogiannakis, George, and William J. Dally. "Router designs for elastic buffer on-chip networks." High Performance Computing Networking, Storage and Analysis, Proceedings of the Conference on. IEEE, 2009.

[13] Agarwal, Niket, et al. "GARNET: A detailed on-chip network model inside a full-system simulator." Performance Analysis of Systems and Software, 2009. ISPASS 2009. IEEE International Symposium on. IEEE, 2009.

[14] Moscibroda, Thomas, and Onur Mutlu. "A case for bufferless routing in on-chip networks." ACM SIGARCH Computer Architecture News. Vol. 37. No. 3. ACM, 2009.

[15] Kodi, Avinash Karanth, et al. "Adaptive inter-router links for low-power, area-efficient and reliable Network-on-Chip (NoC) architectures." Proceedings of the 2009 Asia and South Pacific Design Automation Conference. IEEE Press, 2009.

[16] Hayenga, Mitchell, Natalie Enright Jerger, and Mikko Lipasti. "Scarab: A single cycle adaptive routing and bufferless network." Proceedings of the 42nd annual IEEE/ACM international symposium on microarchitecture. ACM, 2009.

[17] A. T. Tran and B. M. Baas, "RoShaQ: High-performance on-chip router with shared queues," in IEEE Intern. Conf. on Computer Design, Oct. 2011, pp. 232–238.

[18] Ben-Itzhak, Yaniv, Eitan Zahavi, and Israel Cidon. "HNOCS: modular open-source simulator for heterogeneous NoCs." Embedded Computer Systems (SAMOS), 2012 International Conference on. IEEE, 2012.

[19] Kim, Hanjoon, Yonggon Kim, and John Kim. "Clumsy Flow Control for High-Throughput Bufferless On-Chip Networks." (2012): 1-1.

[20] Ding, Hui, et al. "Configuring algorithm for reconfigurable Network-on-Chip architecture." Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on. IEEE, 2012.

[21] Demiriz, Ayhan, Nader Bagherzadeh, and Abdulaziz Alhussein. "Cpnoc: On using constraint programming in design of network-on-chip architecture."Parallel, Distributed and Network-Based Processing (PDP), 2013 21st Euromicro International Conference on. IEEE, 2013.

[22] Comparative Performance Evaluation of Power and Area Network on Chip (NoC) Architectures

[23] Wang, Chifeng, and Nader Bagherzadeh. "Design and evaluation of a high throughput qos-aware and congestion-aware router architecture for network-on-chip." Microprocessors and Microsystems 38.4 (2014): 304-315.

[24] Kahng, Andrew, Bo Lin, and Siddhartha Nath. "ORION3. 0: A Comprehensive NoC Router Estimation Tool."

**Jawwad Latif** received his B.S from COMSATS, Islamabad, University and M.S from UET, Taxila, Pakistan. His research interests include router architectures, topologies and fault tolerant techniques in Network-on-Chip.

**Sadia Azam** received her B.S from COMSATS, Wah cantt, University and M.S from UET, Taxila, Pakistan. His research interests include router architectures and routing techniques in Network-on-Chip.

**Hassan Nazeer Chauhdry** is persuing his PhD. at Polito Turino Italy. He received his MSc. from Technical Univesity Darmstadt, Germany in ICE(2012-2014) and MSc. Electrical Engg., UET Taxila, Pakistan (2007-2009) respectively. He did his computer engineering (2003-2006) from UET Taxila. He worked at UET Taxila as lecturer(2006-2009) and Asst. Prof. (2010-2012). His current area of research includes SOC/ASIC solution for signal processing and communication based applications.

**Tahir Muhammad** is a Phd student in EE at University of Engineering Technology, Taxila, Pakistan. He completed his MSc in Computer Engineering from University of Engineering and Technology Taxila, Pakistan in 2010. He did his B.A.Sc in Electrical Engineering from University of Ottawa, Canada. His areas of expertise are Digital Design and Network on Chip Architectures.