ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

Evaluation of Effect of Packet Injection Rate and Routing Algorithm on Network-on-Chip Performance

Mostafa Haghi[1], M. Asha Rani[2]

P.G. Student, Department of Electronics and Telecommunication, Jawaherlal Nehru Technological university, Hyderabad ,Andhra Pradesh, India[1]

Professor, Department of Electronics and Telecommunication, Jawaherlal Nehru Technological University, Hyderabad, Andhra Pradesh, India[2]

Abstract: In system on chip (SOC) with NOC interconnection paradigm, PIR (Packet Injection Rate) and Routing algorithm are two critical factors that affect the efficiency of SOC communication. Its observed that NOC may saturate at certain rate of injection packets. The effect of PIR on three routing algorithms in switch-based NOC with limited number of IP blocks is studied in this paper. The chosen algorithms for this work are FULLY ADAPTIVE, XY and WEST FIRST, Mesh topology is applied with 16 cores (N*N style). Performance evaluation is conducted based on flit-accurate and open source SystemC simulator.
Eventually the critical points in saturation area are determined for above algorithms and throughput, power consumption and delay are compared for mentioned routing algorithms.

Keywords: packet injection rate, xy, fully adaptive, west first, network on chip

I. INTRODUCTION

System on chip (SOC) in forthcoming billion transistors era because of power density restriction and technology improvement will involve the integration of numerous heterogeneous semiconductor source blocks[1][2]. A source block can be a processor core, memory, an FPGA, a custom hardware block or any other intellectual property (IP)[4].System on chip interconnection play a crucial role to achieve the target performance[5]. Traditionally system on chips utilize topologies based on shared buses, Dally and Towles proposed replacing dedicated design specific wires with general purpose(packet–switched) network[3] ,hence it was the beginning of Network on Chip(NOC) era. Nowadays the NOC is the backbone of SOC design and is progressing toward Wireless NOC paradigm [6]. Routing algorithm is a key factor of network-on-chip which affects the efficiency of NOC communication [7].The routing algorithm is defined as the path taken by a packet between the source and the destination IPS. According to where the decision about routing path is taken, it is possible to classify the routing into category of Source and distributed routing [8]. In Source routing, the whole path is decided on source router, it means

the entire path that the packet is supposed to pass is known before communication while in Distributed routing each router receives the packet from previous IP and decide the next direction and send it[12]. According to how a path is defined to transmit packet to next IP, routing can be classified as Deterministic or Adaptive [8]. In deterministic routing algorithm, packets are routed without considering network's state but in Adaptive, routing uses information about network's state (e.g. channel load information and buffer's size) to make a decision. Deterministic algorithms are

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

widely used due to easy implementation. In Adaptive routing each router has congestion information of its surrounding neighborhood [12]. Here three kind of algorithm means fully adaptive, partially adaptive and deterministic are taken to experiment.

1. XY Algorithm

Wang Zhang and Ligang Hou[9] proposed Classic XY(Static XY OR XY) routing algorithm which is one type of distributed deterministic routing algorithms. they used 2-Dimesion mesh topology and router. To implement classic routing algorithm, router(or IP) is identified by its coordinate (x, y).According to XY routing algorithm the current router address (Cx,Cy)has to be compared with the destination router address (Dx,Dy) of the packet, that has been stored in the header flit. Flits must be routed to the core port of the destination router when the (Cx,Cy) address of the current router is equal to the (Dx,Dy) address. If this is not the case, first the X axis is concerned. therefore the Dx address is compared to the Cx (horizontal) address. Flits will be routed to the East port when Cx<Dx, to West when Cx>Dx and if Cx=Dx the header flit is already horizontally aligned. When the X axis destination router address is true, the Dy (vertical) address is compared to the Cy address. Flits will be routed to South when Cy<Dy, to North when Cy>Dy. If the chosen port is busy, the header flit as along with all subsequent flits of this packet will be in pending state (blocked). The routing request for this packet will remain active until a connection channel is established in future execution of the procedure in this router.

2. WEST FIRST Algorithm

In this algorithm all 90-degree turns to the west are prohibited, a packet in order to travel west has to start out in this direction. This requirement suggests the west-first routing algorithm: route a packet first west, if necessary, and then adaptively south, east, and north [10].

3. FULLY ADAPTIVE Algorithm

The policy of fully adaptive routing algorithm is always using a route that is not congested. The distance is not a matter for algorithm, it means, the algorithm does not care although the path that is taken between sender and receiver IPs is not the shortest direction. Typically an adaptive routing algorithm sets alternative congestion free routes to order of superiority. The shortest route is the best one. [8]

4. PIR

Usually the rate that packets are injected into the network by a node is termed as a packet injection rate pir()(packet/cycle/IP). Pir is restricted between 0 and 1(0<Pir<=1), for instance when pir=.3 it means in each IP sends 3 packets every 10 cycles [15].

5. Acceptable Latency

The packet latency or delay refers to the time spent from the header flit is injected into the network by the source IP to the tail flit is accepted by the destination node[14].

Actually the packet latency can be broken into two parts:

The transferring latency and the waiting time[15].

The transferring latency is the time spent to forward the packet and is irrelevant of network status. However the waiting time which is the time taken to wait some other packet to pass, depends on network's status completely. In case of congestion, the packet latency grows up to hundreds of cycles rapidly.

According to [14] it takes two cycles to deliver a flit from one router to its neighbor router.

In case of waiting time :

Waiting time=2* packet size (flits). (1)

Where in this study the size of each packet is 6flits, and the max numbers of hopes between any two distant IPs is 7(4*4 IP blocks), by calculation we realize that the acceptable time latency is restricted to 26 cycles.

II. RELATED WORK

The flow control is a very important factor that determines the efficiency of Network- on –Chip (NOC). In recent years many researchers have been working on it, especially when the Packet Injection Rate is around saturation point, delay

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

is very sensitive to it, that is the reason the PIR has a obvious effect on NOC performance. Lately a lot of papers about flow control has been presented in literature. Here, we introduce some of works that are related to this work.

P. Avasare et al. [16] proposed a centralized end -to- end traffic management algorithm for packet- switched best-effort NOC. In already mentioned model two independent NOCs are required, one is partitioned the data NOC on which the data packets are transmitted , the other in fact controls the NOC where the control information are collected. The control information are queried by a control OS periodically. The algorithm that the OS is adopted to carry out the traffic management bases on a message injection rate control mechanism. In detail , the message injection rate is adjusted according to the amount of the blocked messages and throughput. The centralized OS can be the bottleneck of communication.

I. Nousias et al. [17] presented an adaptive rate control for NOC based on wormhole switching with virtual channels (VC). When contention arises, the receiving IP(node) asks each source node to adjust their transferring rate accordingly. The requesting node could be the intermediate router requires not necessarily to be the destination of the packet. In this proposal the detail is not described.

III. EXPERIMENTAL RESULTS

Three routing algorithms are considered in this paper, Fully adaptive, xy and west first. The simulation results along with tables are presented for these three, from Fig.1 to 5 and Table1 to 3 repectively.

The traffic that is used here is uniform, in which each IP block randomly sends packet to any other IPs.

This traffic model is oftenly used in many simulations.

As in Fig.1 is seen when pir is incremented from 0.01up to 0.03 total received packets are increased and consequently the throughput and power consumption are increased expectedly ,delay also is increased which is not desired but still is controlable. This policy is continued up to PIR=0.03 , when it exeeds 0.03 an unusual behavior is observed, throughput is degraded rapidly with power consumption ,in the other hand the rate of delay is increased this behavior is observed up to PIR=0.04, in next step, when the PIR is increased, means PIR=0.05 , an interesting result is raised up,the simulator doesn't work and no results are out from PIR=0.05 onwards for fully adaptive routing algorithm. This can not be a saturation case, it occures because the buffer size is limited to 8 in this study and since in fully adaptive routing always a not congested route has to be selected, with increasing the number of received packets, buffer is filled and there is no more space to receive new packets, this is the reason that the number of received packets is 0(zero). It obviously means the throughput is zero and the NOC in fully adaptive routing algorithm is not useable for pir>0.04 .

We can observe that the max throughput is obtained at PIR=0.03, although the time latency also is in high , but its still acceptable.

In contrast with fully adaptive algorithm, in XY routing algorithm situation is different. In fact in this routing algorithm when the PIR exeeds a certain value, the saturation is occured.

Simulation is started with PIR=0.01, as the PIR is Incremented by each step(each step in this work is concerned to 0.01) throughput ,power consumption and delay are increased.This is continued up to PIR=0.04, at this packet injection rate, all Parameters are in MAX level, but the rate of time latency has increased sharper. At PIR=0.05, as is shown in Fig.2 still all parameters are Increasing but two points clearly are observable. First the rate of inremention for throughput and power consumption has been reduced. Second point is that the number of delay cycles has been increased rapidly and is out of Control , for instance at PIR=0.05 the latency is equal to 2645, and when PIR is increased 1 step(PIR=0.06) it will be 10777,this amount of time latency is not acceptable, this is the sign of being in saturation domain because in mentioned area delay is very sensetive to PIR. If we keep increasing the packet injection rate it will be observed that according to Fig.3 from PIR=0.05 onward the saturation is occurred. In saturation state throughput will be variated between 28.1 and 28.7. Power consumption also in this region is rolling between 4.13 and 4.4.

Since the delay in saturation area is very sensetive to PIR, the time latency is rapidly increasing, but after pir>0.3 the rate of delay incremention is decreased and sensetivity degree is reduced, with this number of delay cycle, discussing about the throughput is meaningless.

In WEST FIRST routing algorithm the policy is very similar to XY algorithm But obtained information are deffrent. Simulation is begun with PIR=0.01, up to PIR=0.04 the discussed parameters are increased,with a almost fixed rate,As

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

is observed in Fig.5,at PIR=0.05 time latency is increased with a sharp rate and is out of control. (54.53 at PIR=0.04 to 6240.3 at PIR=0.05).

At PIR=0.05 again saturation state is happened. Number of received packets, power consumption and averagethroughput are almost fixed or with a little changes. In this case the throughput in saturation area is swing between 25.2 and 25.7 and power consumption is varying between 3.95 and 3.99.

A remarkable point is that,the delay incremention rate is reduced after PIR=0.3 in a same way with XY routing algorithm, and from PIR=0.5 onwards the time latency is almost fixed.

It means that the sensetivity to PIR is decreased.

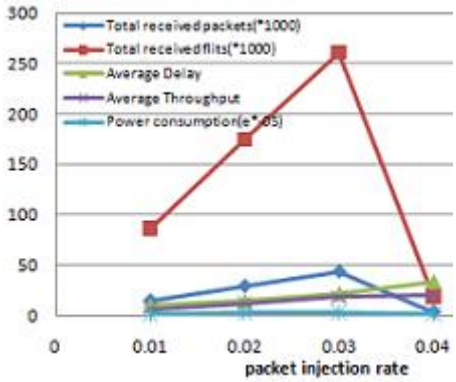

Fig. 1 Power consumption,throughput and delay under fully adaptive routing

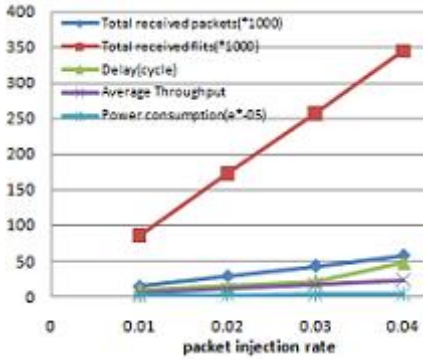

Fig. 2 Power consumption, throughput and delay under xy routing

A. Delay

While study of time latency, we observe that the simulation results are almost same for three routing algorithms within 0.01<=PIR<=0.03, but for PIR>0.03 the rate of latency increment is increased for WEST FIRST routing faster, in contrast with other two algorithms. XY routing algorithm has the better latency than Fully ADAPTIVE, and fully adaptive has the most number of delay cycles in this study.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

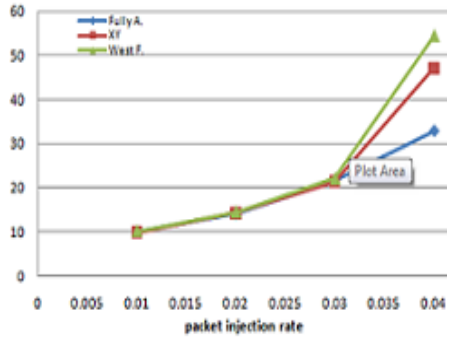

Fig. 6 Delay under Fully A., XY, West F. routing algorithms

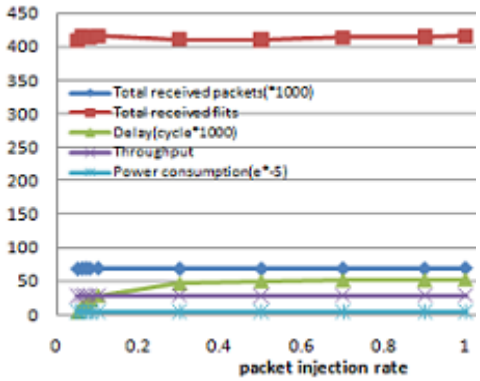

Fig. 3 Power consumption, throughput and delay in saturation state under xy routing

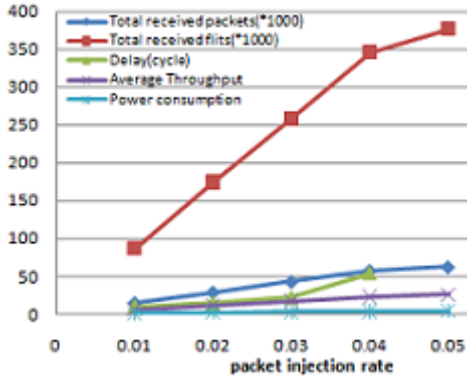

Fig. 4 Power consumption, throughput and delay Under west first routing

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

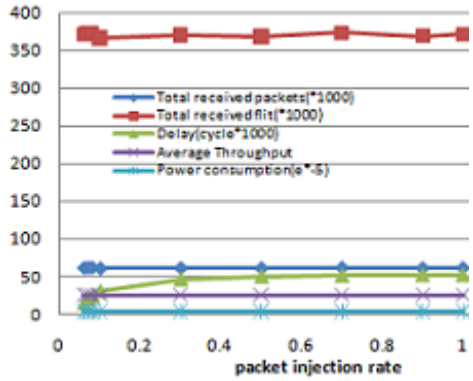

Fig. 5 Power consumption, throughput and delay
State under west first routing

In Fig.7 we can see clearly that for PIR>0.04 the rate of latency increment is decreased and slowly is getting close to saturated surface.

One more point is that, the saturation peak for XY routing algorithm is higher than WEST FIRST routing algorithm,therefore the sensetivity of west first to PIR is lower. As was mentioned before the simulator at pir>0.04 has no response for fully adaptive routing algorithm.

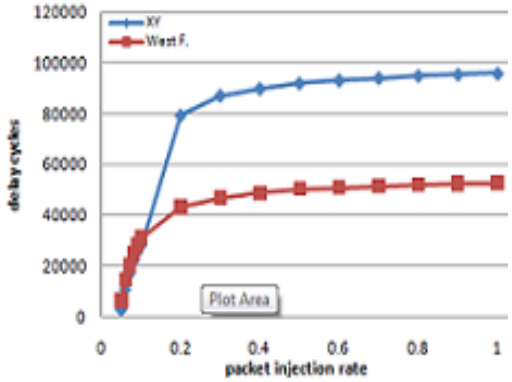

Fig. 7 Full diagram of delay under West F. and XY routing algorithmIn both before saturation and in saturation state

B. Power consumption

Here is shown in Fig.8, at 0.01<=PIR<= 0.03 almost all three algorithm have the same power consumption, but the XY routing is slightly better. When the PIR crosses the 0.04 it starts entering to saturation area,the rate of power consumption increment is decreased and the level of west first routing in saturation area is restricted to lower surface.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

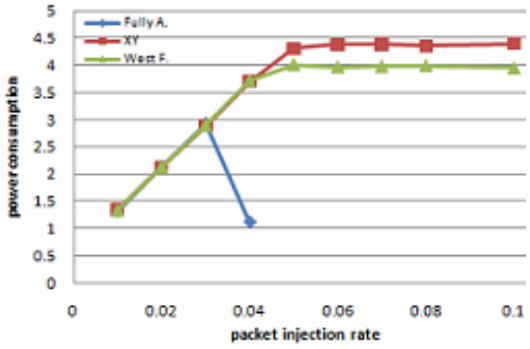

Fig.8 Power consumption under three routing algorithms

C. Throughput

According to Fig.9 The best throughput result in range of 0.01 <=PIR<=0.03 belongs to fully adaptive routing algorithm. When it crosses the PIR=0.03 in fully adaptive routing algorithm throughput fell down rapidly and in next step its stopped working in NoC,Between XY routing and WEST FIRST ,the second case has the better throughput before saturation state,means at 0.03<=PIR<=0.04. In saturation state the throughput of XY is fixed in highr level Than WEST FIRST and swing between 28.1 and 28.6, while wWEST FIRST routing algorithm is varying between 25.6 and 25.9.

High throughput while having a unacceptable latency is meaningless.

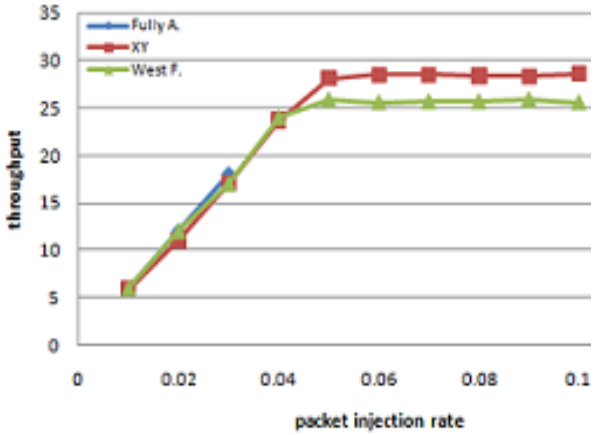

Fig. 9 Throughput under three routing algorithms

Table.1. Time latency for F.A, XY and W.F under different PIR

PIR	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08
F.A	9.93	14.05	21.49	32.95	…	…	…	…
XY	9.89	14.30	21.35	47.16	2645	10777	16693.4	21739
W.F	10.11	14.44	22.02	54.53	6244.3	14570.5	20174	24697

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

Table.2. Throughput for F.A, XY and W.F under different PIR

PIR	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.3	0.4	0.5
F.A	6	12	18	1.2	…	…	…	…	…	…	…	…	…	…
XY	6	11.9	17.7	23.7	28.1	28.5	28.5	28.4	28.3	28.6	28.6	28.3	28.5	28.2
W.F	6	12	17	23	25	25.6	25.7	25.7	25.6	25.2	25.6	25.5	25.4	25.4

Table.3. Power consumption(*e-5) for F.A, XY and W.F under different PIR

PIR	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.2	0.3	0.4	0.5
F.A	1.31	2.14	2.93	1.13	…	…	…	…	…	…	…	…	…	…
XY	1.34	2.13	2.89	3.70	4.31	4.38	4.39	4.36	4.37	4.40	4.39	4.36	4.39	4.35
W.F	1.34	2.14	2.90	3.72	4.01	3.97	3.98	3.99	3.96	3.95	3.98	3.98	3.95	3.95

IV. CONCLUSION

At any given condition, the NOCs saturates at some injection rates, we have shown that in saturate area delay is very sensetive to packet injection rate. Beyond those points,serious congestion occures with increasing pir, and the average time latency increases sharply, on the other hand, congestion should be control through flow control and must not exeeds a certain point. In this paper effect of congested on fully adaptive routing algorithm has been studied and the crucial point has been assigned.

In this work we have determined the critical saturation points for three routing algorithms in 4*4 IP blocks with mesh topology and the behavior of the NOC for mentiond routing algortihms in range of 0.01< pir <=1. Finally we have compared time latency, power consumption and throughput for those three routing algorithms.

REFERENCES

[1] Benini, L.. De Micheli, G., "Networks on chips: anew SOC paradigm" Computer. Volume: **35** Issue: I , Jan. **2002,** pp.: **70 -78**

[2] Saastamoinen, **1.** et al, "Interconnect IP node for Future System.on-Chip Designs", Proceedings of the First IEEE InternationalWorkshop on Electronic Design, Test and Applications, **2002,** pp.**1 I6 -I** 20.

[3] Dally, W.J. and Towles. B.. "Route Packets, Not Wires: On-Chip Interconnection Networks", DAC **2001,** June **18-22,2001,**Las Vegas. Nevada, **USA**

[4] N. Nikitin and J. Cortadella. A performance analytical model for Network-on-Chip with constant service time routers. In *Computer-Aided Design-Digest of Technical Papers, 2009.ICCAD 2009. IEEE/ACM International Conference on*, pages 571–578. IEEE, 2009.

[5] E. Krimer, M. Erez, I. Keslassy, A. Kolodny, and I. Walter. Packet-level static timing analysis for NoCs. In *Networks-on-Chip, 2009. NoCS 2009. 3rd ACM/IEEE International Symposium on*, page 88. IEEE, 2009.

[6] J. Hu and R. Marculescu. Energy-and performance-aware mapping for regular NoC architectures. *IEEE Transactions on computer-aided design of integrated circuits and systems*,24(4):551–562, 2005.

[7] D. Bertozzi, A. Jalabert, S. Murali, R. Tamhankar, S. Stergiou, L. Benini, and G. De Micheli. NoC synthesis flow for customized domain specific multiprocessor systems-onchip.*IEEE Transactions on Parallel and Distributed Systems*,16(2):113–129, 2005.

[8] N. Jiang et al. Indirect adaptive routing on large scale interconnection networks. In ISCA, June. 2009.

[9] Wang Zhang, Ligang Hou, Jinhui Wang, Shuqin Geng,Wuchen Wu, 2006, Comparison Research between XY and Odd-Even Routing Algorithm of a 2-Dimension 3X3 Mesh Topology Network-on-Chip.

[10] Ni. L, Gui, Y., Moore, **S.,** "Performance Evaluation of Switch-Based Wormhole Networks", IEEE Transactions on Parallel and Distributed Systems, Volume: 8 Issue: **5.** May **1997,** pp. **462-474**

[11] W.J. Dally, H. Aoki: *Deadlock-Free Adaptive Routing inMulticomputer NetworksUsing Virtual Channels*. IEEE transactions on Parallel and Distributed

Systems, 1993, Volume 4, Issue 4, pages: 466–475.

[12] M. Dehyadgari, M. Nickray, A. Afzali-kusha, Z. Navabi: *Evaluation of Pseudo Adaptive XY Routing Using an Object OrientedModel for NOC*. The

17th International Conference on Microelectronics, 13–15 December 2005.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2014

[16]L. S. Peh and W. J. Dally, "Flit-reservation flow control," in Proceeding of the 3rd International Symposium on High-Performance Computer Architecture, 2000, pp.73-84

[14] J. Hu and R. Marculescu, "Application specific buffer space allocation for networks-on-chip router design," in Proceeding of international Conference on Computer Aided Design, 2004,pp. 354-361.

[15] A. Ganguly, K. Chang, P. P. Pande, B. Belzer and A. Nojeh, "Performance evaluation of wireless networks on chip architectures," in Proceedings of the 10th International Symposium on Quality of Electronic Design, pp. 350-355,2009.

[16]P. Avasere , V. Nollet, J. Y. Mignolet, D. Verkest," Centralized end-to-end flow control in a best-effort network-on-chip,"in proceedings of the 5th ACM International Conference on Embedded Software, 2005,pp.17-20.

[17]I. Nousias and T. Arslan," Wormhole routing with virtual channels using adaptive rate control for network on chip(NOC)," in proceedings of the 1th NASA/ESA Conference on adaptive Hardware and Systems,2005,pp.493-498.