ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Indian Journal of Science and Technology, *Vol 9(29), DOI: 10.17485/ijst/2016/v9i29/91707, August 2016*

Low Latency Noc with Dynamic Priority based Matrix Arbiter

J. Arjana*

School of Computing, SASTRA University, Thanjavur - 613401, Tamil Nadu, India; arjanitha@gmail.com

Abstract

The quest for the improvement of processing power and efficiency is spawning research for many core systems. Network on chip (Noc) is evolving as an eminent way in replacing shared buses for better design and reusability. The packet switch fabric posses a high dominant problem which gives rise to high latency and communication uncertainty. Packet requirements can be detected and dispatched from different directions based on the priorities so that the packets pass through the router in a congestion free manner. Network on chip replaces the shared bus system and routing is performed in a multi hop basis. Router architectures have been proposed to reduce the average network delay. However communication uncertainty becomes more critical to system performance. Pipelining stage can further increase the throughput as much as possible. A priority arbitration technique is used in this paper to reduce the average latency by Dynamic priority based matrix arbiter in the pipelining stage. If many packets want to get access to the same output channel, the role of router is to decide which packet should be delivered to the next router. The scheduling algorithm is implemented with the matrix arbitration technique in the pipelining stage which increases the speed of communication. With the help of the tool Xilinx ISE 14.2 the parameters of throughput and latency is analyzed.

Keywords: Arbitration, Dynamic Priority (DP), Matrix arbiter, Network on chip (Noc), Pipelining

1. Introduction

More number of cores integrated on a particular chip increases as, the technology scales down. Due to the increase in the number of modules in the System on chip the bus based will prevent the systems to meet the performance which will be needed by many computer programs. Buses provide the interconnection to be more critical. Buses may not provide the needed frequency, latency, power consumption for intensive parallel communication. The best solution is the use of the embedded switching network, called Network on chip, which interconnect the many IP modules in the System on chip as shown in Figure 1. Network on chip provides a promising interconnect[1] infrastructure for future. In Network on chip packets will be routed in a hop basis. Network on chip provides the rapid evolution for the growing wire delay and increase in arbitration.

Network on chip brings the benefits of scalability and reusability and also inherits the problem in the network. Due to multi-hop packet transmission, the latency in packet-switched network is large in comparison with the bus based systems. Network on chip design provides a useful way in the reduction of latency. Moreover Network on chip provides a distributive way without any central control unit to collect the overall network information that leads in the inability of dealing with the completion for shared resources. In multi-core and many-core architectures packet switch interconnect fabric has been introduced[2]. Packet switch interconnect provides high throughput and scalability. It is capable of sharing the resources on the traffic flows in on-chip networks. Network on chip function as data buffering and arbitration. Packets which belong to different traffic flows will compete for the resources. In order to optimize the wiring and switching resources flits are grouped as a flow control. This communication uncertainty leads to a problem. To reduce the communication uncertainty, various router architectures has been developed. Many researchers

*Author for correspondence

focused on reducing the transmission time of Network on chip. Most of the work is focused on reducing the average network latency and has ignored the latency distribution which is important to the overall system performance of parallel program. By reducing the variance of latency[3,4] it can improve the global synchronization time of the whole system. Virtual channel flow control employs packet switched networks to implement chip wide interconnect network. Priority based technique[5]can be used to adjust packets at the needed runtime and without the knowledge of the application type. Due to long time stalling in the network packet transmission can be speeded. For a continuous success in resource contention it can be slowed down. The former work has been done by using adaptive routing. Adaptive routing technique will maintain the average latency level. In this work we aim to reduce the average latency of on-chip networks by using matrix arbiter scheduling algorithm.

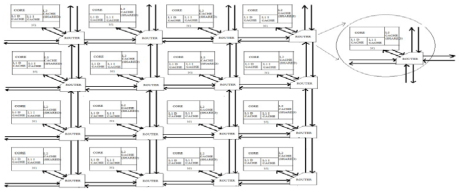


Figure 1. Noc based CMP.

2. Combining Age based and Deadline based Arbiter

The latency variance of the packet should be reduced by not getting them run as fast as possible. The packets should run slow than too run too fast a mechanism is should be used to hurry up the packets when it lags behind the other packet. Priority of the packet[6] should able to upgrade and downgrade. In such case a priority generator is to be needed in which it combines both the age based and deadline based arbiter. In the age based arbiter the priority of the packets will be given for the highest request time. Suppose two or more packets need to reach the same destination from the source the packet which takes more time to reach the destination, the priority will be given[7]. The function of the age based arbiter is to make the packet first run slow then fast by assigning higher priority to the aging packet. The age based arbiter can be of ascending priority as the approximate type. In the Deadline based arbiter the priority will be given for the longest distance transmission[8]. When two or more packets want to reach the same destination the packet which takes the long distance to transmit the priority will be given to it. The function of deadline based arbiter is to make the packet run fast then slow because higher priority is far from the destination and lower one when it approach the destination[9]. The deadline based arbiter can be of descending priority.

3. Router Architecture with Priority Generator

This router architecture includes the priority generator which combines both the age based and deadline based arbiter. The flits in the router read a flit in the Virtual channel FIFO buffer[10]. In the header flit the RC unit gives the output and stores the information in the Virtual channel state register. The Virtual channel request gets passed through the output port priority unit. Function of priority generator is to collect the entire request and the priority will be generated at respective clock cycle. The highest one will be sent to the Virtual channel allocator. The flit which obtains a free Virtual channel gets passed through the Switch Allocator stage at the next cycle while the other will be stalled for one cycle and Request the Virtual channel at the next cycle. For every clock cycle the priority will be updated as shown in Figure 2.

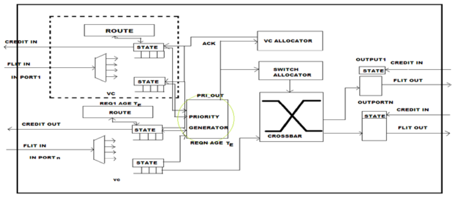


Figure 2. Router architecture with priority generator.

4. Proposed Method

4.1 Router Architecture

In the proposed system pipelining mechanism is used in which dynamic priority based matrix arbiter is implemented. The pipelining concept is included in the router architecture which has router, ejector, injector, switch, and crossbar[11]. In the first stage the router checks for the arrival of the packets that transfers a data with a minimum distance from source to destination. In the ejector stage it removes the incoming flit that arrives from the network. The third stage is the injector stage where it injects the flit that comes from the core and goes to the next stage. The next stage is the switch which connects the incoming packets to a set of output ports which then crosses the final stage of crossbar where is output is delivered as shown in Figure 3.

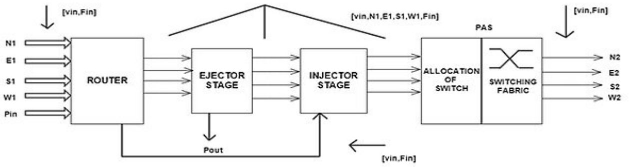


Figure 3. Pipelining stages.

4.2 Dynamic Priority based Matrix Arbiter

The main objective is to develop a dynamic priority with matrix arbitration technique which is independent of deadlocks and is better efficient. According to the traffic load of previous router dynamic priority will be assigned. If many packets want to get access for the same output channel the router is in charge to decide which packet to deliver to the next router, so the dynamic priority based matrix arbitration will be implemented[12]. If all the input port want to get access for the same output port matrix arbiter adopts N x N matrix which gets implemented by priority scheme of a triangular array of bits.The input port competing for the same output port sends the generated grant signals to the one which posses the maximum priority in the matrix[13]. The row in the matrix compete input requests and its priorities. The scheduling module will examine the priorities of each input port request. The arbiter gets updated in the scheduling matrix when the highest priority input is served by making the request which gets the last grant. Row and column gets inversed for the less priority for the next round of arbitration. The block diagram of dynamic priority based matrix arbiter is shown in the Figure 4. It requires three signals as to send request, signal for priority and final signal is for the generation of grant signal. Here the request will be sent to all 4 priority and checks for high priority. In case if 3rd position is to be high then it generates a grant signal as 0010 finally go up to last position. The second block of comparator compares the request signal and priority signal moves to the priority reducer router where the priority is used to enable and grant the signal.

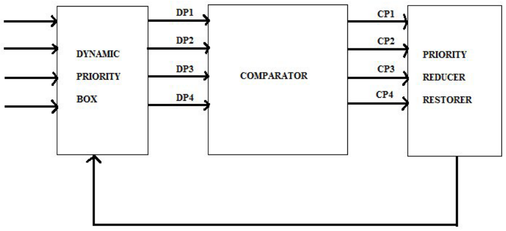
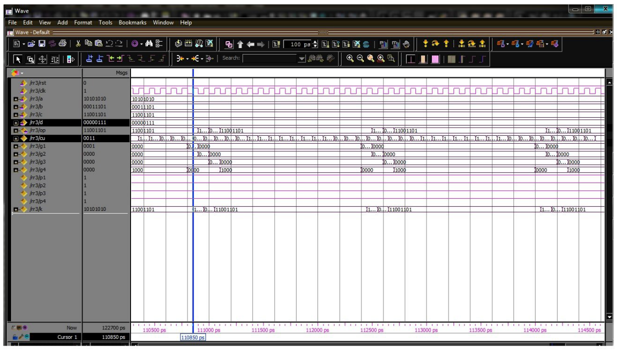


Figure 4. Dynamic priority based matrix arbiter.

5. Experimental Results

The language used for the coding of arbiter is VHDL in which behavioral and structural style of modeling is used. For simulation and synthesis purpose Xilinx ISE 14.2 version is used. RTL view and synthesis report can be easily obtained in this software. In this ISE 14.2 the report are generated for different parameter .For calculating delay check the synthesis report which gives the time period of transmitted port. The Latency and throughput can be found with the help of the synthesis report. The simulation result for the arbitration is shown in Figure 5 which gives priority with respect to the generated grant signal.



```
/rr3/rst    0
/rr3/clk    1
/rr3/a      10101010
/rr3/b      00011101
/rr3/c      11001101
/rr3/d      00000111
/rr3/op     11001101
/rr3/cu     0011
/rr3/g1     0001
/rr3/g2     0000
/rr3/g3     0000
/rr3/g4     0000
/rr3/p1     1
/rr3/p2     1
/rr3/p3     1
/rr3/p4     1
/rr3/k      10101010
Now         122700 ps
Cursor 1    110850 ps
```

Figure 5. Simulation result for arbitration.

6. Conclusion

In this paper we proposed Dynamic priority based Matrix arbiter in the pipelining stage. Hence for designing Network on chip (Noc) the arbitration mechanism like Matrix-arbiter is used. This arbiter detects the input ports with respect to clock cycle and priority of each input port gets adjusted dynamically. Matrix–arbiter reaches through higher clock frequency, few resources is consumed. Matrix–arbiter provides processing of data in a quick manner. By using this method Low latency is achieved and throughput is increased.

7. References

1. Benini L, Micheli GD. Networks on chips: A new SoC paradigm. Computer. 2002 Jan; 35(1):70–8.
2. Beulah HS, Vigneshwaran T, Jasmin M. Survey on energy - Efficient methodologies and architectures of Network-on-Chip. Indian Journal of Science and Technology. 2016 Mar; 9(12):1-8.
3. Mullins R, West A, Moore S. Low-latency virtual-channel routers for on-chip networks. 31st Annual International Symposium on Computer Architecture (ISCA); p. 1-10.
4. Selvaraj G, Kashwan KR. Reconfigurable adaptive routing buffer design for scalable power efficient Network on Chip. Indian Journal of Science and Technology. 2015 Jun; 8(12):1-9.
5. Howard J, Dighe S, Vangal SR. A 48-core IA-32 processor in 45 nm CMOS using on-die message-passing and DVFS for performance and power scaling. IEEE J Solid-State Circuit. 2011 Jan; 46(1):173–83.
6. Yan K, Yang H. A Low Latency Variance NoC Router and Hui Wang Institute of Circuits and Systems. Dordrecht: Springer Science+Business Media; 2012. p. 89-97.
7. Matsutani H, Koibuchi M, Amano H. Prediction router: Yet another low latency on-chip router architecture. IEEE 15th International Symposium on High Performance Computer Architecture (HPCA); 2009 Feb. p. 367–78.
8. Park D, Das R, Nicopoulos C. Design of dynamic priority-based fast path architecture for on-chip interconnects. 15th Annual IEEE Symposium on High-Performance Interconnects; 2007 Aug. p. 15–20.
9. Daneshtalab M, Pedram A, Neishaburi MH, Riazat M, Afzalikusha A, Mohammadi S. Distributing congestions in Noc through a dynamic routing algorithm based on input and output selections. Proceedings of International Conference on VLSI Design; 2007 Jan. p. 546-50.
10. Dally WJ. Virtual-channel flow control. 17th Annual International Symposium on Computer Architecture (ISCA); 1990 May. p. 60–8.
11. Chan C-H, Tsai K-L, Lai F, Tsai S-H. A priority based output arbiter for NoC router. IEEE International Symposium of Circuits and Systems (ISCAS); 2011 May. p. 1928–31.
12. Kim J, Nicopoulos C. A gracefully degrading and energy-efficient modular router architecture for on-chip networks. Proceedings of the 33rd International Symposium on Computer Architecture (ISCA'06); 2006 May. p. 4–15.
13. Fu Z, Ling X. The design and implementation of arbiters for Network-on-chips. International Conference on Industrial and Information Systems; National Key Laboratory of Science and Technology on Communications of UESTC; 2010 Jul. p. 292–5.