Design of NOC Routers

Dinesh Anand Bashkaran, Graduate student,Department of Electrical Engineering ,Rochester Institute of Technology,Rochester,New york
Email: dab8730@g.rit.edu

***Abstract*—As the SOC design methodology undergoes revolutionary improvement, in next few years the development of SOC architecture would be enormous. Considering the main reason behind this is all about having better performance, the interconnect connecting the various components inside the SOC should be equivalent effective to respond to the speed of cores and other components. Having said this, there exists Network on Chips, which is an existing architecture that connects various components. Studies have been already made to get it better with various topology connections, various architecture models, and various router models. It seems to have decent result and helps in keeping the performance up to expectation. In this paper, a case study about already existing architecture is taken and certain logic inside routing is being improved for better performance.**

***Index Terms*—NOC, Routers, SOC, Interconnect, warmhole, switching, data bus.**

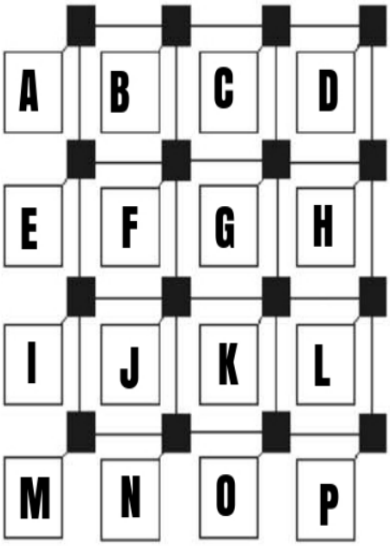
INTRODUCTION:

The size of the transistors decreases time to time, which results in smaller and smaller gates which increase the number of cores inside the chip. The wire that connects the cores remains same irrespective of the technology being used. These wires are called global wires. Global wires run from one corner to the other to carry the data information from chip to chip. It is very certain long wide wire ends up in high capacitance load consuming more power with data latency. Steps have been taken to use buffers to keep the data latency low but then size is being increased. The size reduced by technology is being taken by these buffers which is simply unacceptable. Introduction of Network on chips, which involves breaking of wires and placing routers between them to route the data worked pretty good. Most famous interconnects are ARM AMBA, Wishbone Interconnect and IBM core connect. In this paper NOC architecture using routers is discussed in section 1, the motivation behind this is discussed in section 2 and then the results are discussed in section 3.

I. THEORY:

The work done in this paper is considering CLICHÉ topology. In this topology, the number of routers used is equal to the number of cores inside the chip. Every core is connected to its own routers. Every router has 5 ports to communicate except the one in corner which does not have adjacent core. Figure x, explain how this looks in real time. The ultimate reason behind this is, reduction of area of the wires ends up in low capacitance and avoids the switching of wires that is being used currently. The routers decide the path for the data to flow and keep the wires switching only when required. This way the power consumption is being reduced, besides which the latency of the data being sent is always kept in same clock cycle. This means there is no delay the data sent in cycle T, reaches the destination in cycle T. This technique works well for power consumption reduction and the data latency but then overhead of the router has to be considered. Further down the paper, the area is being addressed.

Fig. 1. NOC for 16 Cores



A. Switching Methodologies:

Switching decides how the routers connect the input and the output from time to time. There are various kinds of switching methodologies.

Circuit Switching; this type of switching involves in reserving the data path before the transmission is about to be held. This path is reserved until the data is being sent. This takes the advantage of entire bandwidth of the wire but then the adjacent cores are held in hold which increases the delay time for them. This gradually reduces the performance of the entire system. No point how fast the cores operate, the hold time for inter-communication is high which results in delay.

Packet Switching; the entire data that needs to be sent is divided into small packets. This method sends the packet which needs to be stored in the router if the path is being used and then send the packet to the destination. This method does not involve the delay but then it increases the size of buffers that has to be placed in the router. Operating at low technology level, providing enormous size to the buffers in Routers is unacceptable.

Warmhole switching; this involves in reducing the data into smaller fixed length control units called FLITS. This reduces the size of the buffer in the routers also this does not involve reserving the entire data path. The FLITS are divided into few categories; the first FLIT is called header flit which stores the information of the destination address which makes the routers to understand the next coming flits should also to transmitted into the same destination. The next FLIT is the body FLIT which stores the actual information; this follows the same path how the header FLIT was sent. It generally makes a pipeline fashion, and then the tail FLIT is sent to acknowledge the data is being sent successfully which is the end of the particular transmission.

B. Performance Matrices:

Throughput is the factor that decides the speed of the circuit. As per our concern throughput is the measure of value that coveys how fast the data is transmitted from origin to destination.

Ways to increase the throughput: - Reduce the distance of travelling ends in having low number of IP blocks travelled. This involves having best algorithm that decides the path to have high throughput and less power consumption - Having the switches operating at high frequency the messages can be passed faster, thought this is trade off to power consumption.

Latency; There is no big difference between this and throughout. This just defines the number of clock cycles it takes for the message to be delivered. Again this can be increased by operating the switches at higher frequency.

power consumption;This is the major issue and we need to heads up to reduce the power consumption as minimum as possible. As of routing, the power consumption is caused due to Switching of routers. Using the buses to transmit data. Using channel to store data (basically they switch on and off to represent 1 and o's). Using registers as channels (also increase area).

Area; The area is split into two inside the routers - Area of the registers or the buffers - The area occupied by the routing logic (multiplexers) So, apart from this we have buffers added to the wire to maintain one clock cycle delay. This is also considered when the total area of the routing architecture is calculated.

Talking about the area, the area occupied by the routing logic is flexible and can be altered with respect to the algorithm being implemented, but the area of buffer is constrain. The ultimate motivation is from this factor where a routing logic is created to have better performance with respect to the power consumption and the latency and also reduce the area of the router. The only chance of reducing the area is by having a routing logic with less overhead, at the same time the logic need to be deadlock and livelock free. This paper takes this as a motivation and developes a method to come up with area effective, high performance routing logic architecture for Routers.

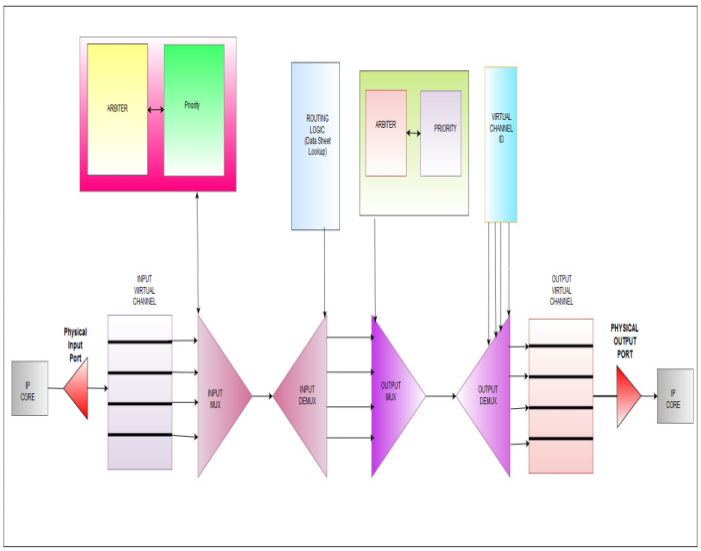
II. Motivation:

The ultimate motivation is to improve the performance of the routers. The routing logic is the CPU of the routers, having them modified to work on better algorithm ends in better performance. Deadlock; is the situation where there is no empty ports for the data to move on and this is usually recovered using reset. Live lock; is the situation where the packets travels away from destination rather than going towards destination. Routing logic is directly responsible for this kind of misbehavior, having a routing that works on X-Y algorithm is not very appropriate. Instead, the routing scheme is already created for shortest path algorithm considering deadlock and live lock and all the routing logic has to do is refer the look up table and route respectively. In this paper, the routing logic is designed accordingly to produce better performance.

III. Design Architecture:

The Figure x, shows the architecture of the Router for the single port. Every router has 5 ports and every IP core has 1 router. Input virtual channels, are basically buffers that gets the data from various cores and stores it in the buffer, this increases the flexibility of physical input port. The input arbiter selects which data from input channel should pass through the input port; this basically works on priority matrix method. Routing Logic, this decides what is the corresponding direction the particular packet has to be transmitted. This routing logic is very critical in avoiding dead lock and live lock. Output arbiter selects what data has to be sent through the output physical port; again it just works on priority.

Fig. 2. NOC Architecture for single Port



A. Look Up Table (LUT):

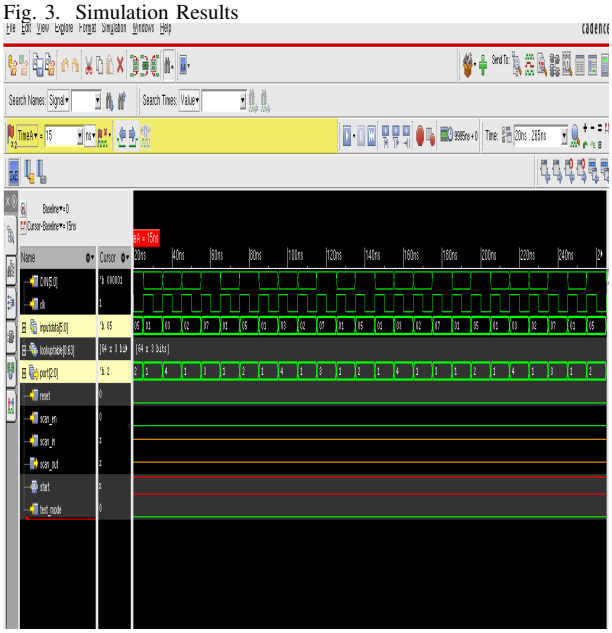
Most critical part of routers is logic that decides how the packet is being circulated which is taken by the routing logic. Other small modules just work on priority. This paper works on modification of Routing technique that works on x-y routing to look up table (LUT) based routing algorithm. The routing logic is designed in such a way; it takes the header

flit and compares it with the look up table which tells the port address. The message is being transmitted to the corresponding port address.

B. Generation of LUT:

The look up lable can be generated using shortest path algorithm such as greedy algorithm. The table should hold the destination address and the corresponding port address. This varies for all the routers as they have different physical location. The generation of LUT table is the key to performance of the entire NOC as this decides how the data communication happens. The LUT is designed in such a way, it reduces the deadlock and livelock issues. Avoidance of using repeated routers and avoidance of using centre routers for major part of communcation helps in reduction of heating density.

Fig. 3. Simulation Results



IV. Results:

As discusses, the Router with routing logic which works on look up table is being created. The routing logic is considered for 64 cores and synthesized. The number of cores is varied across 256, 512, 1024 to view the overhead of the routers. As we discussed, the area of the router is the trade of to the reduction in latency and power consumption. The tradeoff area is calculated and showed as result. Figure 3, shows the results of the simulation performed and figure 4, shows the ports selected with respect to the destination address.

The simulation is performed in cadence vituoso. The RTL is synthesized into hardware and the netlist is generated for this RTL. The figures show the simulation results as dicussed. The basic idea of selecting the ports inside the router using LUT is implemented and simulated. The figure 5, shows the power consumption and the test coverage. The power consumption is reduced by this way of router implementation.

The ultimate idea of this simulation to calculate overhead of the router. The overhead is contributed by the buffer and the

Fig. 4. Routing port with respect to destination

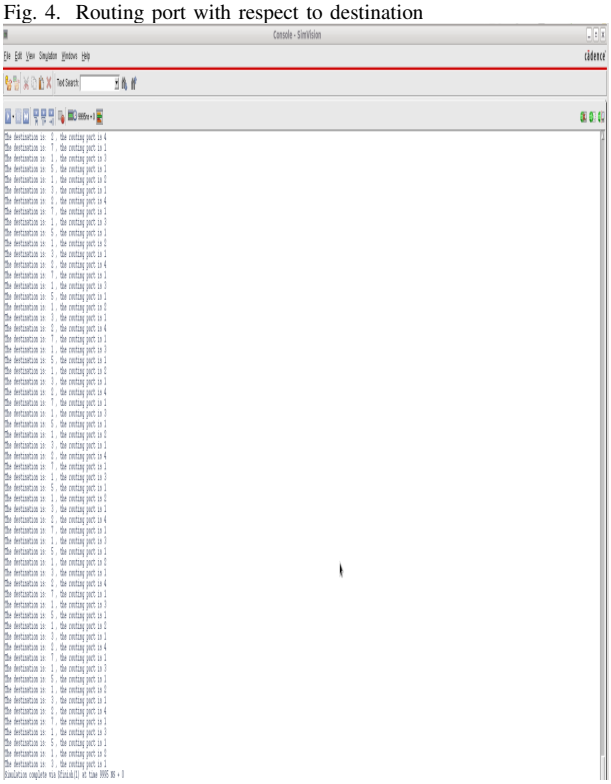
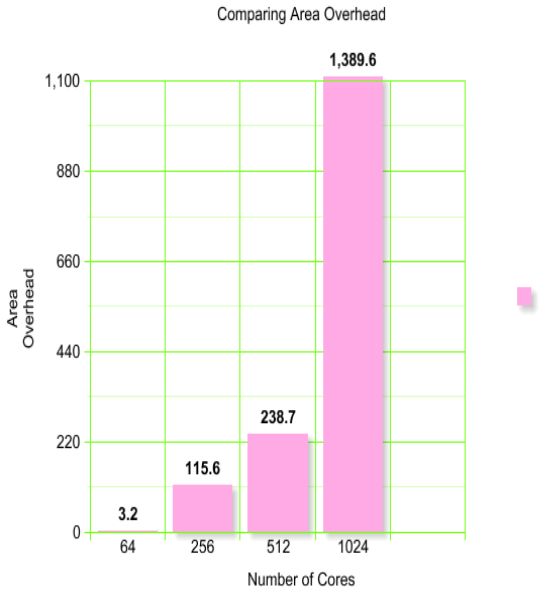


Fig. 5. Comparsion of cores



routing logic. The overhead of the routing logic is calculated here with respect to 64 cores, 256 cores, 512 cores, 1024 cores.

The Area overhead for different core size is shown as the synthesized report. The comparsion graph is shown below which compares the area overhead.

V. Conclusion:

This paper provided new way of implementing the algorithm for the routing technique. Routing logic decides the performance of the NOC. The LUT based routing is implemented

and the area, power consumption and synthesized report are discussed. Also the increase in area overhead of the routers with respect to number of cores is also studied and discussed.

ACKNOWLEGMENT

I would like to thank Dr. Amlan Ganguly who helped me in getting fair amount of work done.

REFERENCES

[1] Pande, Partha Pratim, Cristian Grecu, Michael Jones, Andre Ivanov, and Resve Saleh. "Performance evaluation and design trade-offs for network-on-chip interconnect architectures." IEEE transactions on Computers 54, no. 8 (2005): 1025-1040.

[2] Dally, W. J., & Towles, B. (2001). Route packets, not wires: On-chip interconnection networks. In Design Automation Conference, 2001. Proceedings (pp. 684-689). IEEE.

About the author

Dinesh Anand Bshkaran is a graduate student currently pursuing Masters in Electrical Engineering in Rochester Institute of Technology. His Fields of Interest are Design of Digital Systems, NOC, Embedded Systems, Power Electronics and his hobbies are Gym, Martial Arts.