# A Power Efficient NOC Router Design

M.S.Suraj [#1], D.Muralidharan [#2], R.Muthaiah [#3]

[#] School of Computing, SASTRA University, Thanjavur,
TamilNadu, 613401, India
[1] srjece@gmail.com,
[2] murali@core.sastra.edu
3 sjamuthaiah@core.sastra.edu

*Abstract*—**In this work, we present the NOC router architecture with five port support which utilizes dual crossbar arrangement, the latency which arises due to the dual cross bar architecture is reduced by using predominant routing algorithm. This arrangement is more efficient and reduces about 10 % of device utilization in this work, the use of dynamic voltage scaling (DVS) for links is implemented by predicting the link utilization using the soft computing techniques (ANN) and then shutting the links down based on the threshold which is predicted**

**Keyword- NOC, latency, crossbar, dynamic voltage scaling ,ANN**

## I. INTRODUCTION

The common bus design is the conventional way to connect multiple processors together, where each processor core has a private cache and shares cache on the common bus. But the delay to fetch the data from cache increases significantly with the increase in number of processor cores in the bus [6]. The common bus architecture proves to be inefficient method to connect multiple processors, when the numbers of cores are more. Among many proposed architectures, the packet based Network on Chip (NoC) seems to be the future of multi-core interconnect architecture [2, 5, 7]. In the NoC architecture, the communication between the cores, which is connected to the router, is done through packets. The connection scheme resembles computer networks and hence the term Network On Chip. In order to address the power issues surrounding NOC system large efforts have been taken related to it. Most of the present system focus on shutting the links on and off based on the statistically computed threshold. This paper introduces ANN to compute the threshold value of link based on their utilization values which are dynamic traffic information as a result of dynamic routing.

The organization of this paper is as follows. In Section 2, we introduce the proposed architecture for NoCs. Section 3 results and analysis, and Section 4 concludes the paper giving brief future research directives.
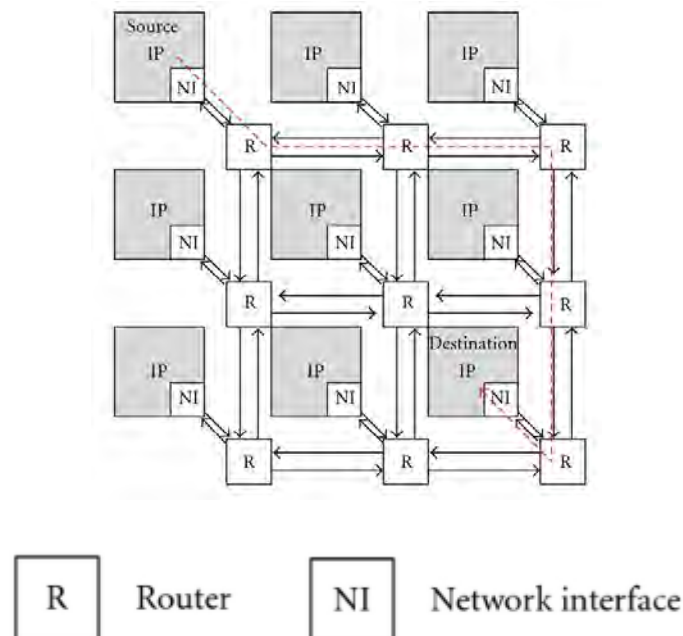


Fig 1 Noc example with connected elements

*A. NoC Architecture*

The high-level architecture of a 4x4 NoC architecture connected in mesh style is shown in Figure 1. The over-all architecture is in many aspects similar to large-scale Networks, where its fundamental components consist of network interfaces, routing nodes, and communication links. On-chip global network communication is supported by a set of interconnected routing nodes that are spread across the chip. Associated with each routing node is a network interface by which node elements (IP blocks) are connected, enabling access to the on-chip network. An NoC architecture is flexible so that it can support different network topologies based on system application requirements.

*B. NoC Router*

Generally, a NoC router has five input and output ports, each of which is for local processing element (PE) and four directions: North, South, West, and East. Each router also has five components: Routing Computation (RC) Unit, Virtual Channel Allocator (VA), Switch Allocator (SA), Flit Buffers (BUF), and Crossbar. When the header flit arrives at the internal flit buffer, the RC unit sends incoming flits to one of physical channels. The Virtual Channel Allocation unit receives the credit information from the neighbouring routers, arbitrates all the header flits which access the same VCs, and then select one of them according to the arbitration policy. Therefore, this header flit can set up the path where the following data and tail flits can traverse this route successfully. The transmitting router sends the control information to the receiving router, and receiving router may update VC ID at the internal buffer with this control information. Switch Allocation (SA) unit arbitrates the waiting flit in all VCs accessing the crossbar and allow only one flit to get crossbar permission. The SA operation is based on the VA stage since the flit data in the buffer comes from the previous router in the route. The flit data pass over the crossbar and thus can arrive at the destination node

*C Artificial neural network*

ANNs are used to determine strong and hidden nonlinearities because of this property they are used in a large application areas even if it is corrupted by noise in the data provided [9]. To name a few applications they are used in forecasting mechanism in stocks and in branch prediction mechanism in computer architecture, as forecasting mechanisms in stocks [9], and in several other prediction applications.
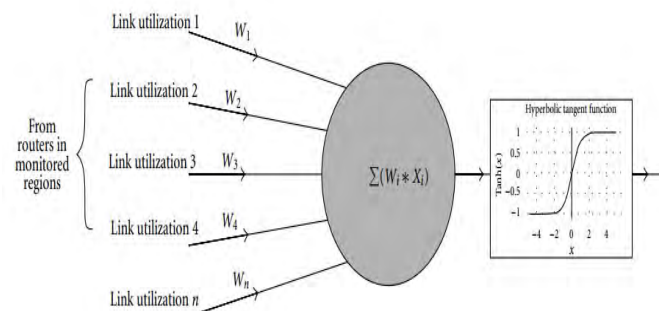


Fig 2 Neuron computation

## II . PROPOSED METHOD

A. *Router Architecture*

The configurable router consists of five ports one for each direction and one to connect the processing element. The Local port connects the processing element with router; other ports provide support and are connected according network topology. Architecture with multiple small crossbar switches provides the same functionality as a single large crossbar, but occupies less area with some increase in router latency. Consequently, the switch used in the configurable router for this paper consists of a dual-crossbar arrangement where two 3X3 crossbars are used instead of a full 5X5 crossbar, as illustrated in Figure 3. The two subcomponents are identified as the first crossbar and the second crossbar because of certain differences in their connections and operations Node elements pass the packets to the router through the local port which in turn goes to the crossbar-one. According to the network topology in which the router is used the packets gets switched directly to West/East output ports or they get passed to second crossbar to be switched to South/North output ports. Packets that enter the South/North port can be passed directly to the node elements but whereas for packets entering through the West/East ports is first switched to the second crossbar from where they are routed to the node element.
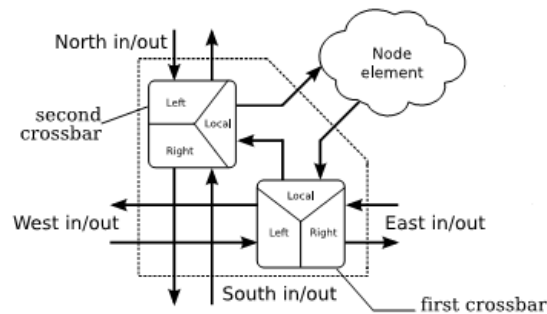
Fig 3 Router with dual cross bar

The switching technique employed by the configurable router is the store-and-forward (SAF) scheme, where the smallest unit of transfer in the network is a packet. Store-and-forward switching requires a packet to be received in it's entirely at each router. A packet is forwarded only when there is space for the complete packet in the receiving router. This scheme is less complex compared to the other switching techniques, but does require appropriate buffer storage. The complexity of SAF switching is further reduced by making the width of the buffer storage and the on-chip interconnect equal to the packet size, thereby avoiding the need to transfer a packet in segments. In this manner, the requirement for full packet reception is easily met. For the intended applications, the SAF scheme provides a convenient basis for the configurable router to form a network infrastructure, with a trade-off involving reduced complexity for potentially higher buffer storage requirements

B. *Switch module*

To connect the input and the output in all possible ways the switch module in each 3X3 crossbar consists of three 3-to-1 multiplexers. This configuration can provide a maximum of three concurrent connections in one crossbar, and a total of six concurrent connections in a dual-crossbar arrangement. Five of the six connections provide support for the bidirectional ports of the overall router, and the remaining connection is used for switching from the first crossbar to the second crossbar .A full 5X5 crossbar is constructed using five 5-to-1 multiplexers in order to provide the same level of overall router connectivity as the dual 3X3 crossbar architecture. The larger 5-to-1 multiplexers do, however, consume significantly more on-chip resources individually than 3-to-1 multiplexers [4], hence five larger multiplexers in the full crossbar require more chip area than six smaller multiplexers in the dual-crossbar architecture.

C. *Routing logic*

When traditional routing algorithms are used in dual crossbar router architecture they tend to increase the latency of the design, in order to minimize this predominant [1] algorithm is used. The used routing algorithm has the advantage of both deterministic and adaptive routing techniques. Based on the traffic characteristics of the network the best path is selected between the source and destination and this path is fixed only for the current flow. The algorithm has of two stages

1. Network setup.
2. Transmission flow

1. Network setup

When a new data is available to be routed to the destination from source then predominant routing algorithm starts its network phase. In this phase two copies of a PATHEXPLORER (PE) flit is generated from the source node or the starting node and it is sent to each outgoing link which has a minimal value, on to the destination. From the minimum cost shortest paths from source to destination Predominant Routing finds the best one. Only two copies of PE flit Arrive at the destination from which the destination node picks the one with the least cost and discards the other Each PE flit has 1) source id, 2) destination id, 3) path details 4) cost. Then the destination node sends the PE flit back to the source node through the chosen path

2. Transmission Flow

After finding the best route ,after receiving the PE flit from the destination node, the source node extracts the routing information from the received flit and prepares to send subsequent data flits through the routing path .The source node reverses the routing path and resends the flits.

D Threshold computation

The threshold for turning the links on and off is computed based on the traffic information which are fed into the ANN as inputs. ANN then predicts the threshold value which is used to turn off the links temporary for short duration of time say for hundred clock cycles during which the status of the link remains unchanged, after this period again the threshold value is computed and the status of the link might change depending on the current link usage, during the whole process the function of the network is maintained with the help of adaptive routing algorithm. The clock cycles for which the link is turned off or on is computed in a trial and error bases if the link was off for a period less than hundred then no significant change in the power level was detected if it was greater than hundred it was unnecessary.

### III. ANALYSIS AND RESULTS

In order to compare the performance of the proposed design, a simple 5 port router, with dual crossbar architecture virtual channel and ANN, without virtual channel and ANN and with dual crossbar architecture with ANN. The entire design was synthesized using Xilinx for vertex- 5 family for XC5VLX50T device
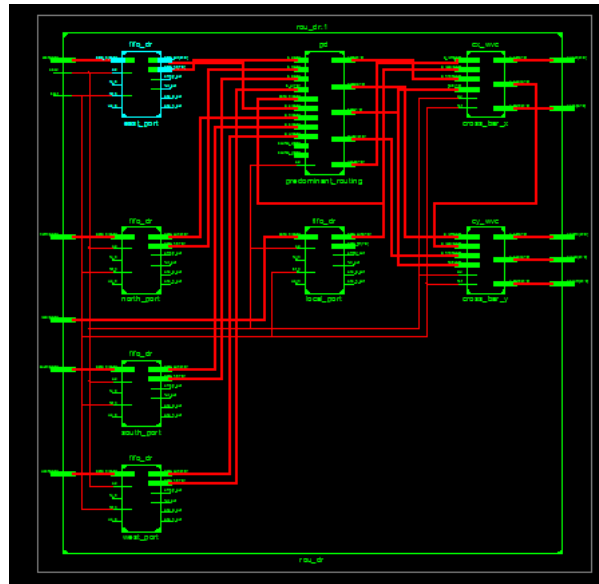


Fig 4 RTL view of dual cross bar router with dynamic routing

Fig 4 gives the RTL view of the entire router from with five port architecture and having  input FIFO in which the incoming data are stored and they are routed with dynamic routing logic – predominant routing which then passes to the dual cross bar architecture .

TABLE I
POWER REPORT

| POWER(mW) | RC-WVC | RC-VC | RC-DR |
|---|---|---|---|
| QUIESCENT(mW) | 473.19 | 449.31 | 449.18 |
| DYNAMIC(mW) | 24.01 | 38.98 | 24.01 |
| TOTAL(mW) | 497.20 | 488.29 | 473.19 |

RC –WVC – cross bar router without virtual channel with xy routing
RC –VC – cross bar router with virtual channel with xy routing

RC–D R – cross bar router with dynamic routing

Power consumption is another important parameter in a NoC router design. Dynamic power consumption is determined by switching activity, leakage power consumption is proportional to the total area. Therefore, leakage power consumption in buffer module is dominant in the total leakage power consumption From the above table it is clear that proposed design reduces both quiescent power and dynamic power

## IV CONCLUSION

In this work, a generic router with virtual channel having dual cross bar architecture is re-examined through implementation using a hardware description language called VHDL. XILINX-ISE is used to synthesize RTL (Register Transfer Level) code and then area and power consumption is extracted from generated code. With the results it is proved that the dynamic power consumption of the proposed router decreases and when used in a mesh network the total power consumption decreases about 10% in comparison with the actual five port router.

## REFERENCES

[1] A. Asad, M. Seyrafi, A. Ehsani Zonouz, M. Seyrafi, M. Soryani, M. Fathy," A Predominant Routing for On-Chip Networks", IDT 2009, pp. 1-6, Riyadh 15-17 Nov. 2009. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] T. Bjerregaard and S. Mahadevan. A Survey of Research and Practices of Network-on-chip. In ACM Computing Surveys, Vol. 38.

[3] J. Duato, S. Yalamanchili, and N. Lionel. Interconnection Network: An Engineering Approach. Morgan Kaufmann Publishers Inc, 2002.

[4] J. Duato, S. Yalamanchili, and L. Ni. Interconnection Networks - An Engi-neering Approach. Morgan Kaufmann, 2003.

[5] N. Agrawal, T. Krishna, Li-Shiuan Peh, and N.K.Jha. GARNET: A detailed on chip network model inside a full-system simulator. In Proc. International Symposium on Performance Analysis of Systems and Software, April 2009.

[6] A. Avakian, J. Nafziger, A. Panda, R. Vemuri. A Reconfigurable Architecture for Multicore Systems. In Proc. IEEE international symposium on Parallel and Distributed Processing, Workshops and Phd Forum. April 2010.

[7] B. Luca, D. Micheli, Giovanni. Network on Chips: Technology and Tools. Morgan Kaufmann, 2006.

[8] R. Pau, "Implementation of a configurable router for embedded network-on-chip support in FPGAs ", EWCAS-TAISA 2008

[9] R. Schalkoff,Artificial Neural Networks, McGrow-Hill, 1997.

[10] Ahmed Hemani, Axel Jantsch and Shashi Kumar, "Network on a Chip: An architecture for billion transistor era"