

Network on Chip Architecture and Routing Techniques: A survey

Amandeep Kaushal¹, Sarbdeep Singh²

¹(M.tech ,E.C.E, Chandigarh Group of Colleges, Mohali, India)

²(Asst. Prof., E.C.E, Chandigarh Group of Colleges, Mohali, India)

ABSTRACT: The processor designing and development was designed to perform various complex logical information exchange and processing operations in a variety of resolutions. They mainly rely on concurrent and sync, both that of the software and hardware to enhance the productivity and performance. With the high speed growth approaching multi-billion transistor integration era, some of the main problems which are symbolized by all gate lengths in the range of 60-90 nm, will be from non-scalable delays generated by wire. All similar problems may be solved by using Network on Chip (NOC) systems. In the presented paper, we have summarized research papers and contributions in NOC area. With advancement in the technology in the on chip communication, faster interaction between devices is becoming vital. Network on Chip (NOC) can be one of the solutions for faster on chip communication. For efficient link between devices of NOC, routers are needed. This paper also reviews implementation of routing techniques. The use of routing gives higher throughput as required for dealing with complexity of modern systems. It is mainly focused on the routing design parameters on both system level including traffic pattern, network topology and routing algorithm, and architecture level including arbitration algorithm.

Index Term – NOC, VLSI, EMI, SOC, CDMA, VCT, QNOC, SPIN.

I. INTRODUCTION

High level of integration in systems with different types of applications is done, where each having its own traffic characteristics. Since the early days of VLSI, using buses is becoming less desirable, especially with the ever growing complexity of single-die multiprocessor systems. As a consequence, the main feature of NOC is the use of networking technology to establish data exchange within the chip. All links in NOC can be simultaneously used for data transmission, which provides a high level of parallelism and helps to replace the typical communication architectures like shared buses or point-to-point dedicated wires [1]. Apart from throughput, NOC platform is scalable and has the potential to keep up with the pace of technology advances. On a chip with billions of transistors, it is not possible to send signals across the chip in real-time bounds [1]. If the SOC is synchronized using global clock signal, the circuit will be exposed to EMI [2]. The SOCs are not power efficient systems. It's not easy for us to synchronize these clock trees because of clock skew problems [3]. As compared to synchronized NOC designs, asynchronous NOC designs are modular and are free suffer from issues like the clock skew, higher power consumption and EMI. Designing efficient circuits, while managing clock arrival time are very complicated in the case of an asynchronous system. There isn't much support given by the EDA industry for asynchronous systems. Thus, researchers tend to hybridize the ideas of synchronous and asynchronous designs. Like in any other network, router is the most important component for the design of communication back bone of a NOC system. In a packet switched network, the functionality of the router is to forward an incoming packet to the destination resource if it is directly connected to it, or to forward the packet to another router connected to it.

II. ARCHITECTURE

From the communication perspective, various topologies for NOC architecture exist. These are as follows: mesh, torus, ring and irregular interconnection networks [4]. Researchers have exploited all these NOC topologies for their NOC implementations. Use of star-based network on chip that communicates using principle of CDMA (Code Division Multiple Access) [5]; Researchers suggested that 2-D mesh for NOC will be highly efficient in terms of latency, energy consumption and easiness of implementation, with comparison to other topologies. The Octagon NOC in [6] is an example of a regular NOC topology.

III. ROUTER ARCHITECTURE

NOC architectures are based upon the packet-switched networks concept. This has led to efficient principles of designing the routers for NOC [7]. Assuming the routers for mesh topologies has only 4 inputs and 4 outputs from/to other routers. Routers are used for implementing various functionalities – ranging from simple

switching to intelligent routing network. As all embedded systems are in a constrained area and power consumption, but still require high data rates, routers with high bandwidth are needed to be designed with hardware usage in mind.

1. SWITCHING TECHNIQUES

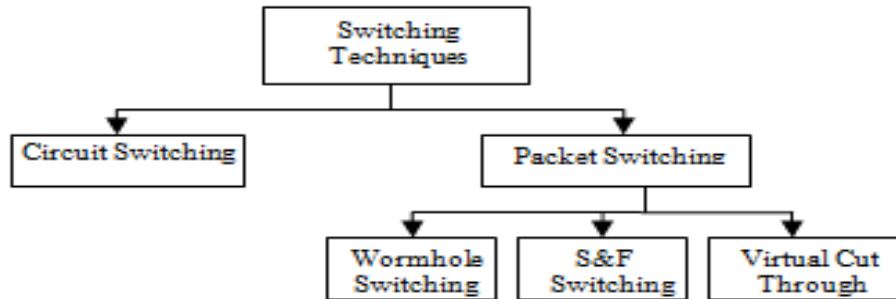


Figure.1 Switching Techniques

Switching techniques are based on network design need. Circuit switched networks preserve usually physical path before they transmitting the data packets, while the packet switched NOC networks transmit packets without using reservation of the entire path. Packet switched networks can be further classified as Wormhole, Store and Forward (S&F), Virtual Cut Switching (VCT) networks (see Figure 1). In Wormhole switching, the header flit experiences latency in flow. Other flits belonging to the same packet will simply follow the path taken by the flit. In case the header flit is blocked then entire set of packet is blocked. It won't require any buffering of the set packets. Therefore, the concerned size of the chip design will drastically be reduced. However, major drawback of this switching technique is the high latency. Therefore, no effort is needed to divide a packet into flits. This will reduce the overheads, as it won't require circuits such as flit builder, flit decoder, stripper and sequencer. Thus, such type of switching techniques requires large amount of buffer memory at every node.

IV. VIRTUAL CHANNEL

The design of a virtual channel (VC) is an important aspect of NOC. Virtual channel splits the single channel into two, providing two paths for packet sets to be routed. There can be 2 to 8 channels. The use of VCs is for reducing the network latency at the cost of area, power use, and production cost of NOC use. However, various other advantages are offered by VCs. Network deadlocks/live-locks: Since VCs provide many output paths per channel there is less probability that the network will suffer from deadlocks; as the network live-lock probability is eliminated (these deadlock and live-lock are different from the architectural deadlock and live-lock, which are due to violations caused in inter-process communication). In today's technology the wire costs are almost similar to the gates. It's likely that in the future the cost of wires dominate a whole lot. Thus, it's important to use the wires very effectively, to reduce cost of a system. A virtual channel gives an alternative route for data traffic, and thus it uses wires more efficiently and effectively for transmission. Therefore, we can reduce wire width on system. For example, we may choose 32-bit instead of using 64-bit. Therefore, the cost of wires and the whole system will be reduced.

V. BUFFER CAPACITY

A high buffer capacity and larger number of virtual channels in the buffering will reduce network contention, resulting in reduced latency. However, buffers are area occupying, and their use needs careful study and optimization. Zimmer and Bolotin both proposed a very simple implementation of the buffer architecture for NOC [8]. The Proteo implementation of buffer system architecture was been described in [9]. Gupta studied the main switching between buffer-size and channel bandwidth in order to secure constant latency in system. They concluded that by overall increase the channel bandwidth will be preferably done to reduce the latency in NOC.

VI. ERROR CORRECTION

The need for implementation of error detection, and correction techniques is not often certain for on-chip system implementations. As proposed a fault tolerant routing protocol for use in the NOC [10]. Bolotin et al. in their implementation system of QNOC [11], [12] argued that all the communication strategies for NOC may be considered to be reliable, while in [13] the proposition of fault tolerant routing as fault tolerant flow control

techniques for NOC architectures are given respectively. As proposed detection and correction schemes for data error on NOC links in [14].

NETWORK INTERFACE

Network interface (NI) controls the packetization and depacketization of all the data traffic, in addition to the conventional interfacing. This functionality may be possible with either hardware or with software. As given comparison of software and hardware implementations for the NI in [15]. They showed that software implementations of NI takes about 47 cycles for completing the packetization/depacketization, while the hardware version takes up to only 2 cycles. Substantial researches were conducted for proposing the right data formats which were needed for various layers used in protocol stack.

QoS

New algorithms have been proposed in the domain to reduce power consumption and space requirements while preserving cost optimization [16]. One of the main concerns of the NOC is to be able to reduce all the latency of an operation. Therefore, various levels of latency programs are offered.

VII. SPECIFIC NOC IMPLEMENTATION

10.1 QNOC

QNOC provides different level of quality and efficiency of service for the all users. The main architecture of QNOC is duly based on a regular mesh topology [18]. It makes the use of wormhole packet routing. Packets are mostly forwarded using the X-Y coordinate routing. Packets are forwarded on the basis of number of credits remaining in the next router. QNOC has identified 4 service levels (SL) similar as that of the NOC communication. The cost functions for QNOC implementation were being calculated based using the estimation and prediction of area occupied by the components. Also there is provision of performance evaluations such as the clock rate, delay, and power consumed under different traffic loads.

10.2 Ethereal NOC

The Ethereal NOC developed by Phillips achieved comparability and predictability in system design and eliminated uncertainties in interconnects, by providing guaranteed throughputs and latency services as in [17], [19]. All queues arranged are 32-bits wide and 8-words deep. Hardware FIFOs are used for implementing these queues. Both router and network interface are implemented in just 0.13 μm technology, and runs at a speed of 500 Megahertz. The NI can deliver the best BW of 16 Gbits/sec to every router in the respective directions. It is all using topology-independent NOC.

10.3 SPIN NOC

The Scalable Programmable Integrated Network-on-chip design is based on fat-tree topologies [16]. This system can have a number of cores depending upon the type of topology. The initiator components are the main traffic generators. The target component sends information as it receives a request. All the components in the system are same as VCI compliant. Such a topology produces a type of non-blocking network with performance that scales gracefully along with the system size. An 8-bit field in header is used for identifying the destination terminals, allowing the network to scale up to all 256 terminals. They all have high level priority when competing with an input buffer to be used as an output channel, which allows reducing of the contention, while minimizing the head-of-line blocking using the freeing of the queues in input buffering system. RSPIN contains 10×10 crossbar, which implements only the selected connections allowed by the routing scheme.

10.4 Other NOCs

A mesh-based NOC which uses the Chip-Level Integration of Communication Heterogeneous Elements (CLICHÉ) [20]; Proteo, which is a flexible-topology NOC as in [21]; A guaranteed-throughput switching for circuit-switched system based NOC, which supports both uni-cast and multicast system as in [22]; A reconfigurable circuit-switched NOC as in [23]; NOCs for the Princeton Smart Camera SOCas given in [24]; A CMOS implementation of a NOC interconnecting various multiple processing units of different system clock frequencies.

10.5 Power Consumption

The leakage current, which is negligible relative to the constantly switching current levels at much larger transistor sizes (1 micron or greater), will be dominating to the current drain with 100 nm technologies. Networks are designed to operate at very low-link utilization in order to meet the worst case scenario in requirements, and thus having a high link capacity which helps to reduce the packet collisions. However, even when NOC links get freed they will still consume power in repeater systems, due to presence of leakage current

small system sizes in NOC. Thus, new techniques will have to consider which will help to reduce the leakage power consumption and will result in a NOC architecture which is more effective.

VIII. ROUTING ALGORITHMS

Routing methods for network decides the path for data transfer to the destination point. The overview of some methods is provided below.

11.1 XY Routing Algorithm

Routing algorithm is used to route the packets from source PE to destination PE. XY routing algorithm is used to route the packet in proposed router designing. It is a type of kind of distributed deterministic routing algorithms. In 2D mesh topology NOC. The XY routing algorithm compares the current router address (Cx, Cy) to the destination router address (Dx, Dy) of the packet, stored in the header flit. Flits must be routed to the core port of the router when the (Cx, Cy) address of the current router is equal to the (Dx, Dy) address. If this last condition 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 well as all subsequent flits of this packet will be blocked. The routing request for this packet will remain active until a connection is established in some future execution of the procedure in this router [26].

11.2 Surrounding XY Routing

Surrounding XY routing (S-XY) has three different routing modes. N-XY (Normal XY) mode works just like the basic XY routing. It routes packets first along x-axis and then along y-axis. Routing stays on NXY mode as long as network is not blocked and routing does not meet inactive routers [11]. SH-XY (Surround horizontal XY) mode is used when the router's left or right neighbour is deactivated. Correspondingly the third mode SV-XY (Surround vertical XY) is used when the upper or lower neighbour of the router is inactive. The SH-XY mode routes packets to the correct column on the grounds of coordinates of the destination. The algorithm bypasses packets around the inactive routers along the shortest possible path. The situation is a little bit different in the SV-XY mode because the packets are already in the right column. Packets can be routed to left or right. Thus the other routers do not send the packets backwards. Surrounding XY routing is used in a DyNOC. It is a method that supports communication between modules which are dynamically placed on a device.

11.3 OE Routing Algorithm

OE routing algorithm is a distributed adaptive routing algorithm which is based on odd-even turn model. It has some restrictions, for avoiding and preventing from deadlock appearance [27]. Odd-even turn model facilitates deadlock-free routing in two-dimensional (2D) meshes with no virtual channels. In this model, a column is called even if its x dimension element is even numerical column. Also, a column is called odd if its x dimension element is an odd number. A turn involves a 90-degree change of travelling direction [27]. A turn is called an ES turn if it involves a change of direction from East to South.

11.4 Switching Methods

It is an important method that can determine connections between input port and output port. The crossbar switch is used in most of the router design for providing full connectivity [28]. There are two basic data switching methods.

11.5 Circuit Switching

In circuit switching routing decision is made when path is set up across the sender and receiver. A dedicated path is established between the sender and receiver which are maintained for the entire duration of transmission [9]. Moreover, links remain occupied even with the absence of data transmission. There is no delay in data flow because of the dedicated path. The major drawback of circuit switching is its limiting scalability.

11.6 Packet Switching

In packet switching, data is broken up into packets. Individual packets take different routes to reach the destination. Each packet includes a header with source, destination and intermediate node address information. The performance can be increases due segmentation of data. There are three types of packet switching: worm hole (WH), store and forward (SAF) and virtual cut through (VCT) switching [29]. In wormhole switching message packets are also pipelined through the network [28]. In SAF switching router should have sufficient buffer space to store the entire packet. Router in every hop must wait to receive the entire packet before forwarding header flit to the neighbouring router [29]. So, the buffer size should be large enough to store entire packet which suffers it from larger latency compared with other technique.

IX. CONCLUSION

In the current technologies the enhancement in the utilization of unused sources instead of inserting new ones can make the NOC an ideal solution for current applications. The NOC mainly concerns of computing and communication, and is very much compatible with the increasing complexity in design and declining system productivity. Researchers will address NOC architectures and hardware-related issues. Still, a vivid approach for modelling, designing and with NOC architecture is missing. Application and feasibility of applications for NOC are some parameters that need to be addressed in more dense detail. We need to research about a low cost, area and power efficient solution for NOC to be effectively implemented in the embedded systems industry application.

REFERENCES

- [1] Fayez Gebali, Haytham Elmiligi, Mohamed Watheq El-Kharashi, "Networks-on-chips: theory and practice", Taylor & Francis Group, LLC, 2009.
- [2] A.S. Kale, Prof. M. A. Gaikwad, "Design and analysis of on chip router for network on chip", International Journal of Computer Trends and Technology- July to Aug Issue 2011.
- [3] A. Louri, J. Wang, "Design of energy-efficient channel buffers withrouter bypassing for network-on-chips (NOCs)", International Journal of Internet Computing, Vol.1, ISS- April 2012.
- [4] J. Dally and B. Towles, Practices of Interconnection Networks, Morgan Kaufmann, 2004.
- [5] P. Pratin Pande, C. Grecu, M. Jones, A. Ivanov, and R. Saleh, "Performance evaluation and design trade-offs for network-on-chip interconnect architectures", IEEE Transactions on Computers, vol. 54, no. 8, pp. 1025-1040, 2005.
- [6] F. Karim A. Nguyen, and S. Dey, "An interconnect architecture for networking systems on chips", IEEE Journal on Micro High Performance Interconnect, vol. 22, issue 5, pp. 36-45, Sept 2002.
- [7] D. Kim, Manho Kim, and G.E. Sobelman, "CDMA-based NOC architecture", Proc. IEEE Conference on Circuits and Systems, vol. 1, pp. 137-140, 2004.
- [8] E. Bolotin, A. Morgenshtein, I. Cidon, R. Ginosar, and A. Kolodny, "Automatic hardware-efficient SOCintegration by QoS Network-on-Chip", Proc. 11th International IEEE Conference on Electronics, Circuits and Systems, pp. 479-482, 2004.
- [9] I. Saastamoinen, M. Alho, and J. Nurmi, "Buffer implementation for Proteo network-on-chip", International IEEE Proceeding on Circuits and Systems, vol. 2, pp. 113-116, May 2003.
- [10] C. Albenes, Zeferino Frederico G. M. E. Santo, Altarniro Amadeu Susin, "ParlS: A parameterizable interconnect switch for Networks-on-Chips", Proc. ACM Conference, pp. 204-209, 2004.
- [11] E. Bolotin, I. Cidon, R. Ginosar, and A. Kolodny, "QNOC: QoS architecture and design process for network on chip", Journal of Systems Architecture, Volume 50, Issue 2-3 (Special Issue on Network on Chip), pp. 105-128, February 2004.
- [12] E. Bolotin, I. Cidon, R. Ginosar and A. Kolodny, "Cost considerations in Network on Chip", Integration: The VLSI Journal, no. 38, 2004, pp. 19-42.
- [13] M. Pirretti, G. M. Link, R. R. Brooks, N. Vijaykrishnan, M. Kandemir, and M. J. Irwin, "Fault tolerant algorithms for network-on-chip interconnect", Proc. IEEE Proceeding on Computer SOCIety, pp. 46-51, February 2004.
- [14] H. Zimmer and A. Jantsch, "A fault model notation and error-control scheme for switch-to-switch buses in a network-on-chip", Proc. First International IEEE/ACM/IFIP Conference on Hardware/Software Codesign and System Synthesis, pp. 188-193, 2003.
- [15] P. Bhojwani and R. Mahapatra, "Interfacing cores with on-chip packet-switched networks", Proc. 16th International IEEE Conference on VLSI Design, pp. 382-387, 2003.
- [16] P. Bhojwani, R. Mahapatra, J. K. Eun, and T. Chen, "A heuristic for peak power constrained design of network-on-chip (NOC) based multimode systems", Proc. IEEE International Conference on VLSI Design, pp. 124-129, 2005.
- [17] E. Rijpkema, K. Goossens, A. Radulescu, J. Dielissen, J. van Meerbergen, P. Wielage, and E. Waterlander, "Trade-offs in the design of a router with both guaranteed and best-effort services for networks on chip", IEE Proc. on Computers and Digital Techniques, vol. 150, Issue 5, pp. 294-302, September 2003.
- [18] T. Bjerregaard and J. Sparso, "Virtual channel designs for guaranteeing bandwidth in asynchronous Network-on-Chip", Proc. of IEEE Norchip Conference, pp. 269 - 272, November 2004.
- [19] K. Goossens, J. Dielissen, and A. Radulescu, "A Ethereal network on chip: Concepts, architectures, and implementations", IEEE Design & Test of Computers, vol. 22, Issue 5, pp. 414-421, September 2005.
- [20] S. Kumar, A. Jantsch, J-P. Soiminen, M. Forsell, M. Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, "A network on chip architecture and design methodology", IEEE Computer, pp. 117-124, 2002.
- [21] I. Saastamoinen, M. Alho, and J. Nurmi, "Buffer implementation for Proteo network-on-chip", International IEEE Proceeding on Circuits and Systems, vol. 2, pp. 113-116, May 2003.
- [22] T. Bjerregaard and J. Sparso, "Virtual channel designs for guaranteeing bandwidth in asynchronous Network-on-Chip", Proc. of IEEE Norchip Conference, pp. 269 - 272, November 2004.
- [23] P. T. Wolkotte, G. J. M. Smit, G. K. Rauwerda, and L. T. Smit, "An energy-efficient reconfigurable circuit-switched network-on-chip", Proc. 19th IEEE International Conference on Parallel and Distributed Processing Symposium, pp. 155-163, 2005.
- [24] J. Xu, W. Wolf, J. Henkel, S. Chakradhar, and T. Lv. "A case study in networks-on-chip design for embedded video", IEEE Proc. on Design Automation and Test in Europe Conference, vol. 2, pp. 770-775, February 2004.
- [25] S. J. Lee, K. Lee, S. J. Song, and H.-J. Yoo, "Packet-switched on-chip interconnection network for system-on-chip applications", IEEE Trans. on Circuits and Systems-II, vol. 52, no. 6, pp. 308-312, 2005.
- [26] Wang Zhang, Ligang Hou, Jinhui Wang, Shuqin Geng, Wuchen Wu, "Comparison Research between XY and Odd-Even Routing Algorithm of a 2-Dimension 3X3 Mesh Topology Network-on-Chip", IEEE Computer SOCIety, 2011.
- [27] Wang Zhang, Ligang Hou, Jinhui Wang, Shuqin Geng, Wuchen Wu, "Comparison Research between XY and Odd-Even Routing Algorithm of a 2-Dimension 3X3 Mesh Topology Network on Chip", IEEE Computer SOCIety, 2009.
- [28] S. S. Mehra, R. Kalsen and R. Sharma, "FPGA based Network-on-Chip Designing Aspects" National Conference on Advanced Computing and Communication Technology, ACCT-10, 828.
- [29] P. B. Domkondwar, Dr. D. S. Chaudhari, "Implementation of Five Port Router Architecture using VHDL", International Journal of Scientific & Engineering Research, Vol.3, Issue 5, May 2012.