Exploring Alternative Topologies for Network-on-Chip Architectures

Shafi Patel[1], Parag Parandkar[2], Sumant Katiyal[3] and **Ankit Agrawal[4]**

Submitted in April 2011; Accepted in August 2011

Abstract - ***With increase in integration density and complexity of the system-on-Chip (SOC), the conventional interconnects are not suitable to fulfill the demands. The application of traditional network technologies in the form of Network-on-Chip is a potential solution. NoC design space has many variables. Selection of a better topology results in lesser complexities and better power-efficiency. In the proposed work, key research area in Network-on-chip design targeting communication infrastructure specially focusing on optimized topology design is worked upon. The simulation is modeled using a conventional network simulator tool packet tracer 5.3, in which by selecting proposed Topology 35.7 % reduction in traversing the longest path is observed.***

Index Terms - NoC, SoC, Routing, Mesh, packet tracer

1. INTRODUCTION

Recent technological development in the field of integrated circuits has enabled designers to accommodate billions of transistors. The level of integration has enhanced computational power enormously. The exponential decrease in the feature size has enabled integration of heterogeneous IP cores on a single chip leading to a new era of integration circuits known as System-on-Chip. However, as the number of components and their performance continue to increase, the design of power, area and performance efficient communication infrastructure is gaining equal importance. The traditional methods of connecting these heterogeneous IP Cores are not meeting the demands of these very complex structures. Furthermore, with technology scaling, traditional global interconnects cause problems like synchronization errors, unpredictable delays and high power consumption. [1] Traditional bus and crossbar based methods to communication become very inefficient, resulting in massive numbers of wires, failed timing closure, increased heat and power consumption, and routing congestion leading to increased die area. The Network–on-Chip approach promises the alternative to traditional bus-based and point-to-point communication structures. The networking methods have been dealing with same kind of problems on traditional computer networks. It indicates that NoC designers can borrow the concept of conventional computer networking with necessary customization to suit demands of SoCs.

The SoCs consists of heterogeneous IP-Cores such as Video processors, Image processors, memory blocks etc. Each of these cores is connected to NoC through a network interface or network adapter module. The NoCs contain a network of routers responsible for end to end delivery of the packets from IP-cores. The communication demands of these IP-cores vary depending on the application running on it. The network interface provides seamless integration of theses IP-Cores and network. Locating the interconnect logic closest to each IP block results in fewer gates, fewer and shorter wires, and a more compact chip floor plan. Having the option to configure each connection's width, and each transaction's dynamic priority assures meeting latency and bandwidth requirements. The routers are connected to each other through links. The origin of NoC has also been viewed as a paradigm shift from computation centric to communication-centric design as well as the implementation of scalable communication structures. The modular architecture of NoC makes chip structure highly scalable and well controlled electric parameters of the modular block improve reliability.

As the network communication latency depends on the characteristics of the target application, computational elements and network characteristics (e.g. network bandwidth and buffer size [2]. First of all the target applications and their associated traffic patterns and bandwidth requirements for each node in the network is determined. This application partitioning and knowledge of overall system architecture significantly impact the network traffic and helps determine the optimal network topology. Optimal network topology creates immense impact of design cost, power and performance and helps designers to choose effective and efficient routing algorithms and flow control scheme to manage incoming traffic.

The design space of a NoC is very large, and includes topology choice (mesh, torus, star, etc.), circuit switched or packet switched, and other parameters (link widths, frequency, etc.). Because the traffic patterns of most SoCs can be known, a custom generated network topology and physical placement of components yields better performance and power than a regular-pattern network [4]. A NoC's buffers and links can consume near 75% of the total NoC power [5], thus there is significant benefit to optimizing buffer size, link length and bandwidth of a NoC design.

Generally speaking, determining the optimal topology to implement any given application does not have a known theoretical solution. Although the synthesis of customized architectures is desirable for improved performance, power consumption and reduced area, altering the regular grid-like structure brings into the picture significant implementation issues, such as floor planning, uneven wire lengths (hence, poorly controlled electrical parameters), etc. Consequently,

[1,2] Chameli Devi School of Engg., Indore, [3]School of Electronics, DAVV, Indore,[4]Techoz Solutions, Indore
E-Mail: [1]shafi.techoz@gmail.com, [2]p_paragp@yahoo.com, [3]sumant578@yahoo.com and [4]ankit.agrawal512@gmail.com

ways to determine efficient topologies that trade-off high-level performance issues against detailed implementation constraints at micro- or nano-scale level need to be developed.

2. BACKGROUND

Network-on-Chip (NoC) is an emerging paradigm using packet switched networks for communications within large VLSI system-on-Chip. NoCs are poised to provide enhanced performance, scalability, modularity, and design productivity as compared with previous communication architectures such as busses and dedicated signal wires. With the emergence of large number of cores in general purpose and system-on-chip (SoC), NoCs are likely to be prevailing on-chip interconnect fabric. [6]

The early work and basic principles of NoC paradigm were outlined in various seminal articles, for example [7-17] and few text books [18-20]. However, the aforementioned sources do not present many implementation examples or conclusions.

Networking concepts from the domains of telecommunication and parallel computer do not apply directly on chip. From a networking perspective, they require adaptation because of the unique nature of VLSI constraints and cost e.g. area and power minimization are essential; buffer space in on-chip switches are limited, latency is very important, etc. At the same time, there are new degrees of freedom available to the network designer, such as the ability to modify the placement of network endpoints. From the view point of VLSI designer, many well understood problems in the real aim of chip development methodology get a new slant when they are formulated for a NoC based system, a new trade-offs need to be comprehended. Therefore, the field offer opportunities for noble solutions in network engineering as well as system architecture, circuit technology, and design automation. [6]

Current complex on-chip systems are also modular, but most often the modules are interconnected by an on-chip bus. The bus is a communication solution inherited from the design of large board- or rack-systems in the 1990's. It has been adapted to the SoC specifics and currently several widely adopted on-chip bus specifications are available [31-34].

While the bus facilitates modularity by defining a standard interface, it has major disadvantages. Firstly, a bus does not structure the global wires and does not keep them short. Bus wires may span the entire chip area and to meet constraints like area and speed the bus layout has to be customized [35]. Long wires also make buses inefficient from an energy point of view [36]. Secondly, a bus offers poor scalability. Increasing the number of modules on-chip only increases the communication demands, but the bus bandwidth stays the same. Therefore, as the systems grow in size with the technology, the bus will become a system bottleneck because of its limited bandwidth.

Recently, network-on-chip (NoC) architectures are emerging as a candidate for the highly scalable, reliable, and modular on-chip communication infrastructure platform [11]. The NoC architecture uses layered protocols and packet-switched networks which consist of on-chip routers, links, and network interfaces on a predefined topology. There have been many architectural and theoretical studies on NoCs such as design methodology [10], [11], topology exploration [21], Quality-of-Service (QoS) guarantee [22], resource management by software [23], and test and verifications [24].

In large-scale SoCs, the power consumption on the communication infrastructure should be minimized for reliable, feasible, and cost-efficient implementations. However, little research has reported on energy- and power-efficient NoCs at a circuit or implementation level, since most of previous works have taken a top-down approach and they did not touch the issues on a physical level, still staying in a high-level analysis. Although a few of them were implemented and verified on the silicon [25], [26],they were only focusing on performance and scalability issues rather than the power-efficiency, which is one of the most crucial issues for the practical application to SoC design.

Network-on-Chip (NoC) architectures employing packet-based communication are being increasingly adopted in System-on-Chip (SoC) designs. In addition to providing high performance, the fault-tolerance and reliability of these networks is becoming a critical issue due to several artifacts of deep sub-micron technologies. Consequently, it is important for a designer to have access to fast methods for evaluating the performance, reliability, and energy-efficiency of an on-chip network. [27]

While on-chip networks have been proposed and studied in the academic literature, to date there have been very few implementations of routed on-chip networks. Dally and Towles [10] proposed a 2D torus network as a replacement for global interconnect. They claim that on-chip network modularity would shorten the design time and reduce the wire routing complexity. On-Chip routed networks have also been proposed for use in SoCs such as in CLICHÉ [12] , in which a 2D mesh network is proposed to interconnect a heterogeneous array of IP blocks.

A performance analysis also shows that dynamic resource allocation leads to the lowest network latencies, while static allocation may be used to meet QoS goals. Combining the power and performance figures then allows an energy-latency product to be calculated to judge the efficiency of each of the network [28].

In his work, Nikolay K. Kavaldjiev , used run-time reconfigurable NoC for streaming DSP applications taking the advantage of a global comunication architecture that avoids limitation by structuring and shortening the global wires. He also proposed an architecture of a virtual channel router, which in contrast to conventional architectures is able to provide predictable communication services and has a lower implementation area and cost than conventional architectures. Dynamic reconfiguration is essential to support the dynamically changing demands of the application domain: the system operates in a constantly changing environment. The user demands change (e.g., starting/terminating applications), the environmental conditions change (e.g., available networks, wireless channel conditions) and the available power budget

also changes (decreasing battery budget or connected to the mains). The set of running applications and tasks in the system adapts dynamically to these changes. The run-time reconfiguration modifies the system communication demands. For example, a new data stream may be needed or some of the old streams may be redirected or replaced. The NoC must be able to handle such dynamically changing traffic conditions. Run-time changes in part of the traffic must be possible without disturbing the rest of the traffic. The network reconfiguration time must be short enough to enable adequate system reaction time and reconfiguration must be transparent to the user. [30]

The major goal of communication-centric design and NoC paradigm is to achieve greater design productivity and performance by handling the increasing parallelism, manufacturing complexity, wiring problems, and reliability. The three critical challenges for NoC according to Owens et al. are: power, latency, and CAD compatibility [17]. The key research areas in Network-on-Chip design can be summarized as [29]:

Communication infrastructure: topology and link optimization, buffer sizing, floorplanning, clock domains, power

Communication paradigm: routing, switching, flow control, quality of service, network interfaces

Benchmarking and traffic characterization for design and runtime optimization

Application mapping: task mapping/scheduling and IP component mapping.

3. METHODOLOGY

Network-on-Chip is a new paradigm for interconnecting today's heterogeneous IP cores based System-on-Chips (SoCs). In SoC's IP Cores are connected to network of routers using network interfaces and network is used for packet switched on-chip communication. Conventional computer design tools i.e. Packet Tracer 5.3 utility from CISCO are used for network design and simulation. It provides a versatile practice and visualization environment for the design, configuration, and troubleshooting of network environments. The work done by us uses same tool to compare two topologies. The 2-D mesh is currently the most popular regular topology used for on-chip networks in tile-based architectures, because it perfectly matches the 2-D silicon surface and is easy to implement. However, a number of limitations have been proved in the open literature, especially for long distance traffic. In this type of topology, every node has a dedicated point to point link to every other node in the network. This means each link carries traffic only between the two nodes it connects.

If N is total no of nodes in network. Number of links to connect these nodes in mesh = N (N-1)/2 Each node should have (N-1) I/O ports as it require connection to every another node.

The advantages are:

1. No traffic problem as there are dedicated links. Robust as failure of one link does not affect the entire system.
2. Security as data travels along a dedicated line.
3. Points to point links make fault identification easy.

The disadvantages are:

1. The hardware is expansive as there is dedicated link for any two nodes and each device should have (N-1) I/O ports.
2. There is mesh of wiring which can be difficult to manage.
3. Installation is complex as each node is connected to every node.

As earlier studies have shown that maximum power is consumed by links and interconnect infrastructure. Reducing interconnects and links will result in lower power consumption but can also affect the performance and reliability negatively. The topology suggested by us reduces the number of links thus resulting into lower power consumption keeping same level of reliability and performance levels.

4. SIMULATION

As shown in figure 1 and figure 2, the number of links in the mesh topology is 24 while in proposed topology the number of links are 20. The number of hops a packets traverses in the longest path is 5 in Figure 1 while 4 in Figure 2. The time taken by a packet to traverse the longest path is 0.014 sec in Mesh topology while 0.009 sec in proposed topology as shown in table 1 and table 2. The percentage reduction in time a packet takes on longest path is 35.7 %.

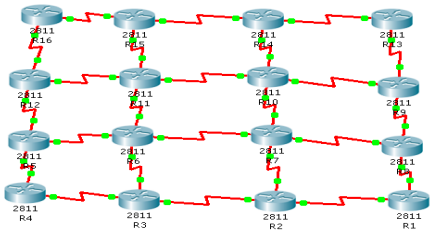

Figure 1: Mesh topology

S.No.	Source Node	Destina-tion Node	Interme-diate Node	Time Elapsed (sec)
1	R1	R16	R8	0.003
2	RI	R16	R9	0.005
3	R1	R16	R13	0.008
4	R1	R16	R14	0.011
5	R1	R16	R15	0.013
6	R1	R16	R16	0.014

Table 1: Result analysis of Mesh topology

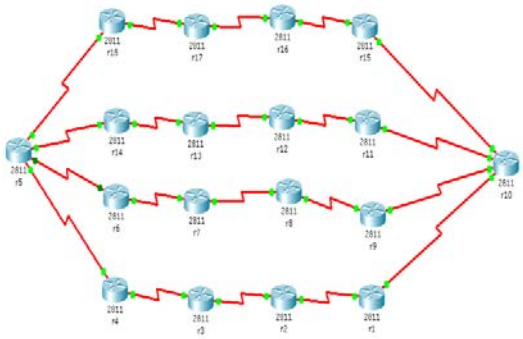

Figure 2: Proposed topology 1

S. No.	Source Node	Destinati-on Node	Inter m-ediate Node	Time Elapsed (sec)
1	R1	R18	R10	0.003
2	RI	R18	R15	0.004
3	R1	R18	R16	0.005
4	R1	R18	R17	0.007
5	R1	R18	R18	0.009

Table 2: Result analysis of proposed topology 1

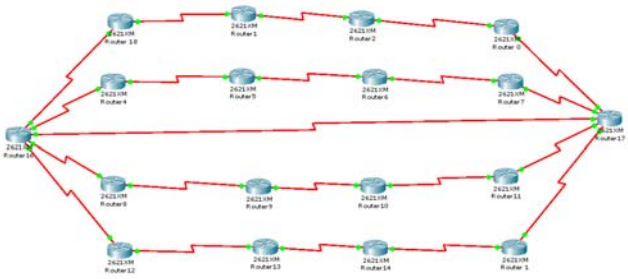

Figure 3: Proposed topology 2

S.No.	Source Node	Destinati-on Node	Interm-ediate Node	Time Elapsed (sec)
1	R1	R18	R17	0.001
2	RI	R18	R16	0.002
3	R1	R18	R18	0.003

Table 3: Result analysis of proposed topology 2

Comparison of proposed topology 2 and 3 shows further improvement in total flight time in traversal of a packet on the longest path. Addition of one link between R17 and R16 reduces the traversal time as well as number of hops.

5. CONCLUSION

The results achieved in terms of time and reduction in number of links displayed here is encouraging and motivates us to take the work further. As discussed earlier the NoC technology can borrow the tools and techniques from conventional computer network technology with required customization. In our future work, we intend to test same on a standard NoC benchmark. The other design parameters on NoC will also be explored.

REFERENCES

[1]. M. Horowitz, R. Ho, and K. Mai, "The future of wires", Proc. IEEE, vol. 89, pp. 490-504, Apr 2001.

[2]. Radu Marchulescu et. al. outstanding research problems in NOC design: system, micro architecture and circuit perspectives IEEE transactions on computer aided design of integrated circuits and systems vol. 8 , No.1,Jan 2009.

[3]. R.Dafali, J-Ph.Diguet and M.Sevaux, "Key Research Issues for Reconfigurable Network-on-Chip", 2008 IEEE International Conference on Reconfigurable Computing and FPGAs

[4]. Murali, S., P. Meloni, F. Angiolini, D. Atienza, S. Carta, L. Benini, G. D. Micheli and L. Raffo, "Design of application-specific networks on chips with floorplan information," in: Proc. of the International Conf. on Computer-Aided Design, 2006.

[5]. Srinivasan, K. and K. S. Chatha, "A low complexit heuristic for design of custom network-on-chip architectures", in: Proceedings of DATE, 2006

[6]. Avinoam Kolodny & Li-Shiuan Peh, IEEE transactions on VLSI systems, vol 17, no. 3 , Mar 2009

[7]. D. Wiklund and D. Liu, "Switched Interconnects for System-on-a-Chip Designs," in IP2000, Oct 2000, pp. 192-198.

[8]. P. Guerrier and A. Greiner, " A Generic Architecture for On-Chip Packet-Switched Interconnections," in DATE March 2000, pp. 250-256.

[9]. A. Hemani, A Jantsch, S. Kumar, J. Oberg, M. Millberg, and D. Lindqvist , "Network-on-Chip: An Architecture for Billion Transistor Era," in Norchip, 2000

[10]. W. Dally and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks," in DAC, June 2001,pp. 684-689

[11]. L. Benini and G.de Micheli, "Networks on chip: A New SoC Paradigm," IEEE computer Vol 35, No. 1 pp. 70-78, Jan 2002

[12]. S. Kumar et al ," A Network-on-Chip Architecture and Design Methodology," in VLSI, Apr. 2002, pp. 105-112

[13]. P. Wielage and k. goossens, " Networks on Silicon: Blessing or Nightmare?," in DSD, Sep 2002, pp. 196-200

[14]. J. P. Soininen , A. Jantsch, M. Forsell A. Pelkonen, J.Kreku and S. Kumar , "Extending Platform-based Design to Network –on-Chip Systems," in VLSI Design, Jan 2003, pp 401-408

[15]. J. Hankel , W. Wolf and S. Chakradhar, " On-Chip Networks: A Scalable Communication-Centric Embedded System Design Paradigm," in VLSI Jan. 2004, pp 845-851.
[16]. U. Y. Ogras, J. Hu, and R. Marculescu, " Key Research Problems in NoC Design: A Hollistic Prespective," in CODES , Sep. 2005, pp. 69-75.
[17]. J. Owens et al , " Research Challenges for On-Chip Interconnection Networks," IEEE Micro, vol 27,no. 5, pp. 96-108, Sep-Oct 2007.
[18]. A. Jantsch and H. Tenhunen , Eds. , "Networks on Chip," Dordrecht , The Netherlends: Kluwer Academic Publishers, 2003.
[19]. J. Nurmi, H. Tenhunen, J. Isoaho, and A. jantsch Eds. "Interconnects – Centric design for Advanced SoC and NoC," Dodrecht , The Netherlends: Kluwer Academic Publishers, 2004.
[20]. "Networks on Chip: Technology and Tools," Morgan Kaufmann, 2006.
[21]. D. Bertozzi et al., "Xpipes: A network-on-chip architecture for gigascale system-on-chip," IEEE Circuits Syst. Mag., vol. 4, no. 2, pp. 18–31, 2004.
[22]. E. Rijpkema et al., "Trade offs in the design of a router with both guaranteed and best-effort services for networks on chip," in Proc. Des., Autom. Test Europe Conf. Mar. 2003, pp. 350–355.
[23]. V. Nollet et al., "Operating-system controlled network on chip," in Proc. Des. Autom. Conf., Jun. 2004, pp. 256–259
[24]. J.-S. Kim et al., "On-chip network based embedded core testing," in Proc. IEEE Int. SoC Conf., Sep. 2004 ,pp.223–226
[25]. S.-J. Lee et al., "An 800 MHz star-connected on-chip network for application to systems on a chip," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2003, pp. 468–469
[26]. M. Taylor et al., "A 16-issue multiple-program-counter microprocessor with point-to-point scalar operand network," in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2003, pp. 170
[27]. Jongman Kim Dongkook Park Nicopoulos, C. ijaykrishnan, N. Das, C.R., "Design and analysis of an NoC architecture from performance, reliability and energy perspective", symposium on architecture for etworking and communication systems, ANCS 2005, pp. 173-182
[28]. Banerjee, A. Wolkotte, P.T. Mullins, R.D. Moore, S.W mit, G.J.M.," An Energy and Performance Exploration of Network-on-Chip Architectures", IEEE transactions on VLSI systems, vol 17, no. 3, March 2009
[29]. Salminen et al. , "Survey of Network-on-Chip Proposals," OCP-IP, Mar 2008
[30]. Nikolay K. Kavaldjiev, "A run-time reconfigurable Network-on-Chip for streaming DSP applications," CTIT Ph.D.-thesis Series No. 06-91
[31]. ARM, "AMBA Specification," Rev. 2.0, 1999, http://www.arm.com/
[32]. IBM, "The CoreConnect™ Bus Architecture," 1999, http://www.ibm.com/
[33]. OCP-IP, "Open Core Protocol Datasheet," 2004, http://www.ocpip.org/
[34]. VSIA, "On-Chip Bus Attributes Specification," Rev. 1, 2001, http://www.vsi.org/
[35]. N. Thepayasuwan and A. Doboli, "Layout conscious bus architecture synthesis for deep submicron systems n chip," in Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE 2004), Paris, France, 2004, pp. 108 -113.
[36]. P. T. Wolkotte, G. J. M. Smit, N. Kavaldjiev, J. E. Becker, and J. Becker, "Energy Model of Networks-on-Chip and a Bus," in Proceedings of the International Symposium on System-on-Chip (SoC'05), Tampere, Finland, 2005, pp. 82-85.