ISSN 2395-1680 (ONLINE)
DOI: 10.21917/ijme.2015.0015

NETWORK ON-CHIP AND ITS RESEARCH CHALLENGES

K. Paramasivam
Department of Electronics and Communication Engineering, Karpagam College of Engineering, India
E-mail:kp_sivam@yahoo.com

Abstract

Networks-On-Chip (NoCs) have been proposed as a promising solution for power, performance demands and scalability of next generation Systems-On-Chip (SOCs) to overcome the several challenges of current SoC with conventional architecture. In this article, NoC, its architecture and features are presented. Further the article is extended with research challenges. Major areas of scope for research are addressed briefly with the view that microelectronic field researchers get benefitted. Performance analysing parameters and simulation tools for NoC are also included. Future SoC design needs lot of innovations and creativity to explore its complete features. Research on NoC is mandatory at this critical juncture.

Keywords:

Network on Chip, System on Chip, Power Dissipation, Research Challenges, NOC Architecture, Network Interface

1. INTRODUCTION

Today's modern lifestyle has been transformed with the development of latest micro electronic technologies like pervasive and ubiquitous computing, ambient intelligence, communication and internet. The developed electronic gadgets are enhancing the way of communication, learning and entertainment. SoC, the latest development in solid state electronics is vital driving force for these advancements. SoC packs complete system in a single integrated chip. The most important features of SoC are smaller in size, faster in speed, larger in capacity, lighter in weight, lesser in power consumption and lesser in cost. Though it is developed in great extent SoC has following challenges to grow further;

1. Signal integrity effects in deep submicron technology such as interconnect delay, crosstalk, inter symbol interference, substrate coupling, etc.
2. Synchronization problem between global and local circuits due to non-proportionate technology scaling in wires and gates.
3. Bus System: The bus system implemented in most of SoCs is dedicated signal wires. The parameters such as bandwidth, clock frequency and power are not improved as expected due to non-proportionate scaling between bus and system size. Further, the bus structure has following limitations.
   a. Limited parallel data transfer capability
   b. Performance degradation with increase in integration technology
   c. Energy inefficient in data transfer
4. Power dissipation: Power dissipation is becoming critical problem in latest VLSI circuits as operating frequency is increased. It is very challenging task to solve the power problem with constrained parameters such as clock frequency, supply voltage, aspect ratio of device, threshold voltage and leakage current.

Several other factors like verification, testing, quick-time-to market, etc. are several drawbacks of SoC design in conventional architecture.

In order to design the SoC effectively, Network on Chip is proposed. NoC is an advanced concept that overcomes the challenges focused by conventional chip architecture with dedicated bus structure between modules and IP cores in SoC. It combines the features of IC technology and communication paradigm to replace the traditional bus structure. NoC connects the IP cores through set of links and routers which are placed in a given network topology.

NoC is a Communication subsystem on an Integrated circuit (commonly called a "chip"), typically between IP cores in a system on a chip. The architecture of SoC are getting communication- bound both from physical wiring and distributed computation point of view. Wiring delays are becoming dominates over gate delays, which favours short links. The overall computation is heterogeneous and localized rather than evenly balanced over the chip for larger Soc. These two factors motivate NoC that brings the techniques developed for macro-scale, multi-hop networks into a chip[1].

Network-On-Chip technology is hastily shifting traditional bus and crossbar approaches for SoC on-chip interconnect. Many terms are being used in the industry. On-Chip Networks, Interconnect Fabrics, Networks-On-Chips and so on. The term “on-chip interconnect” is considered an umbrella name for all approaches. The Network-on-Chip is one specific architecture and is defined as “an on-chip interconnect with decoupled transaction layer, transport layer and physical layer”[1].

The wires in the links of the NoC are shared by many signals. A high level of parallelism is achieved, because all links in the NoC can operate simultaneously on different data packets. Therefore, as the complexity of integrated systems keeps growing, a NoC provides enhanced performance (such as throughput) and scalability in comparison with previous communication architectures (e.g., dedicated point-to-point signal wires, shared buses, or segmented buses with bridges).

NoCs can span synchronous and asynchronous clock domains or use unclocked asynchronous logic. NoC technology applies networking theory and methods to on-chip communication and brings notable improvements over conventional bus and crossbar interconnections. NoC improves the scalability of SoCs, and the power efficiency of complex SoCs compared to other designs.

As the number of IP modules in Systems-on-Chip increases, bus-based interconnection architectures may inhibit these systems to meet the performance required by many applications. For systems with intensive parallel communication requirements buses may not provide the required bandwidth, latency, and

power consumption. A solution for such a communication bottleneck is the use of an embedded switching network, called Network-on-Chip, to interconnect the IP modules in SoCs [2].

Network-on-chip is the most important area of research carries out by many R&D institutions all around the world. The latest contributions on key design issues at different levels of abstraction, namely, design technology, architecture design & optimization, physical link design, performance and power characterization. Applying the networking concept to on chip communication is part of the breakthrough solutions urged by the advances in silicon manufacturing technology [3].

The term NoC is used in research today in a very broad sense ranging from gate level physical implementation, across system layout aspects and applications, to design methodologies and tools [4].

Traditionally, ICs have been designed with dedicated point-to-point connections, with one wire dedicated to each signal. For large designs, in particular, this has several drawbacks from a physical design viewpoint. The wires occupy much of the area of the chip, and in nanometer CMOS technology, interconnects dominate both performance and dynamic power dissipation, as signal propagation in wires across the chip requires multiple clock cycles.

NoC links can reduce the complexity of designing wires for predictable speed, power, noise, reliability, etc., thanks to their regular, well controlled structure. From a system design viewpoint, with the advent of multi-core processor systems, a network is a natural architectural choice. A NoC can provide separation between computation and communication; support modularity and IP reuse via standard interfaces, handle synchronization issues, serve as a platform for system test, and hence, increase engineering productivity.

With NoCs, it is possible to take advantage of part of the technology developed for packet switched networks in the field of communication theory and computer networks, adapting those concepts to the particular constraints of on chip interconnection. NoC can achieve a very high degree of flexibility, using modularity (extensive use of parameterized independent functional blocks) and reconfigurability (functional blocks can be mutually connected in different manners in order to create the particular topology needed for a given application). NoC provides good scope to easily integrate IP cores developed by different people or companies, provided that this module shares a common interface for communication with the external environment.

The basic properties of the NoC paradigm are listed as follows [5]:

1. It separates communication from computation.
2. Avoids global, centralized controller for communication.
3. Allows arbitrary number of terminals.
4. It has a topology that allows the addition of links as the system size grows.
5. Does not utilize long, global wires spanning the whole chip.
6. Allows multiple voltage and frequency domains.
7. It delivers data in-order either naturally or via layered protocol.
8. Offers varying guarantees for transfers.
9. It offers support for system testing.

1.1 NoC ARCHITECTURE

NoC is composed of three main blocks. Network Interfaces (NI), Router and Links is shown in Fig.1. NI makes the logic connection between the IP cores and the network, since each IP have a distinct interface protocol with respect to the network.

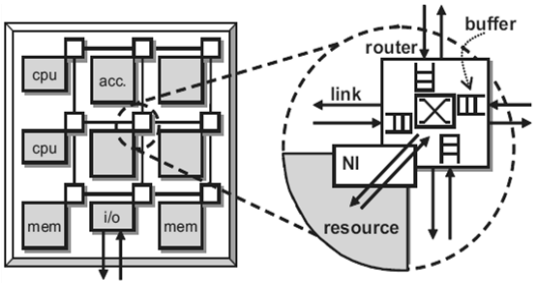

Fig.1. Basic Architecture of NoC [1]

Router is like a "much smarter buffer". It basically receives the packets from the shared links and according to the address informed in each packet, it forwards the packet to the core attached to it or to another shared link. Router is composed of a number of input and output ports, a switching matrix connecting the input ports to the output ports and a local port to access the IP core connected to this router. The buffering policy is the strategy used to store information in the router when there is congestion in the network and a packet cannot be forwarded right away [2].

Links are used to physically connect the nodes and establish the communication. It composed of a set of wires and connects two router in the network. The concept of flits is defined at this level. A packet which is split into smaller data unit is known as flit.

2. RESEARCH CHALLENGES IN NoC

The major goal of communication-centric deign 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 [1].

2.1 CRITICAL ANALYSIS OF NoC

In order to overcome the problems associated with bus-based interconnects, several approaches for networks-on-chip have been proposed, thereby employing reuse of communication resources and providing enhanced reusability and programmability by using standardized and layered communication protocols. A popular approach for NoCs is to employ a regular mesh structure with packet switching, thereby allowing total system bandwidth to increase for each additional network element, since the number of simultaneous communication paths increase. In contrast to the shared bus approaches, NoCs provide greater performance scalability and multiple simultaneous transactions are supported, resulting in more efficient network resource utilization. Furthermore, the

relation between wire length and system size depends on the distribution of area between network elements, and hence, on the regularity of the network; a 2D mesh NoC assumes a constant wire length, independent of system size [6].

2.2 NoC ISSUES AND CHALLENGES

To improve system productivity, it is very important that an architect be able to abstract, represent and address most of the design issues and concerns at a high level of abstraction. System-level design affords one the opportunity to review several different software-hardware architectures that meet the functional specifications equally well, to quickly trade-off among different QoS metrics such as latency, power, cost, size and ease of integration. Similarly, there are several issues related to NoC, such as the nature of the NoC link, link length, serial vs parallel links, bus vs packet-based switching, and leakage current [7].

2.2.1 Serial vs Parallel Link:

The transportation of data packets among various cores in a NoC can be performed by the use of either a serial or a parallel link. Parallel links make use of a buffer-based architecture and can be operated at a relatively lower clock rate in order to reduce power dissipation. However, these parallel links will incur high silicon cost due to inter-wire spacing, shielding and repeaters. This can be minimized up to a certain limit by employing multiple metal layers. On the other hand, serial links allow savings in wire area, minimization in signal interference and noise, and further eradicate the need for having buffers. Serial links offer the advantages of a simpler layout and simpler timing verification. Serial links sometimes suffer from ISI (Inter-symbol Interference) between successive signals while operating at high clock rates. Nevertheless, such drawbacks can be addressed by encoding and with asynchronous communication protocols.

2.2.2 Interconnect Optimization:

Communication in a NoC is based on modules connected via a network of routers with links between the routers that comprise of long interconnects. Thus it is very important to minimize interconnects in order to achieve the required system performance. Timing optimization of global wires is typically performed by insertion of repeaters. It results in a significant increase in cost, area, and power consumption. Recent studies indicate that in the near future, inverters operating as repeaters will use a large portion of chip resources. Thus, there is a need for optimizing power on the NoC. Encoding is an effective way of reducing dynamic power consumption. In order to make NoC architectures more effective, innovative ways will have to be introduced to optimize the power consumed by the on-chip repeaters [7].

2.3 CLASSIFICATION OF RESEARCH AREA

The NoC research area is classified as system level, network adapters, network and link level [4]. The system level research is composed of:

1. Design methodology and abstraction.
2. Architecture domain: system composition, clustering and reconfigurability.
3. Traffic characterization: latency-critical, data-stream and best effort.

The purpose of the network adapter (NA) is to interface the core to the network and make communication services transparently available with a minimum of effort from the core. At this point, the boundary between computation and communication is specified. The research area in network adapters is given below,

1. Encapsulation, service management
2. Sockets: plug and play, IP reuse

The emerging research area in network are listed below

1. Topology: regular vs irregular topologies, switch layout
2. Protocol: routing, switching and control schemes
3. Flow control: deadlock avoidance, virtual channels and buffering.
4. Quality-of-service: service classification and negotiation.
5. Features: error-correction, broadcast, multicast, narrowcast and virtual wires.

The research area in link level is synchronization, reliability and encoding.

The system includesapplications and architecture (cores and network). At this level, most of the network implementation details may still be hidden. Much research done at this level is applicable to large scale SoC design in general.

The NA is the first level which is network aware. The network consists of the routing nodes, links, etc, defining the topology and implementing the protocol and node-to-node flow control.

The network adapter decouples the cores from the network. It handles the end to end flow control, encapsulating the messages or transactions generated by the cores for the routing strategy of the Network. These are broken into packets which contain information about their destination, or connection-oriented streams which do not, but have had a path setup prior to transmission.

The lowest level is the link level. The research in link level regards the node-to-node links. Hence links consist of one or more channels which can be either virtual or physical. Link-level research deals mostly with encoding and synchronization issues [4].

The key issues in NoC research are Noc network architecture, Noc network performance analysis and NoC communication refinement.

2.3.1 Topology:

In network architecture the topology defines how the nodes are interconnected in a network. Numerous topologies including mesh, torus, fat tree, butterflies [8], crossbar, express cube and irregular [9] have been proposed. The most common topologies are 2-D mesh and torus which constitute over 60% of cases. Both have connection between 4 neighbour nodes but torus has wraparound links connecting the nodes on network edges [1].

2.3.2 Switching Techniques:

Switching techniques defines how the data is transmitted from the source node to destination node. In the circuit switching approach, a path is formed from source to destination

prior to transfer by reserving the routers and links. All the data follow that route and path is torn down after the transfer has completed [1].

Packet based switching is one, where all flits of the packet are sent, as the header establishes connection between routers [2]. There is a need to buffer packets in the switch, which implies that switches will occupy more space on the silicon area. The packet based switch is classified as wormhole switching, Virtual cut through switching, and switch & forward switching.

In wormhole switching, the messages are divided into smaller fixed length flow units called flits. A limitation of this method is that messages are not interleaved, only one message can sent over a given physical connection at the same time [9].

Switch and Forward switching technique forwards a packet only when there is enough space available in the receiving buffer to hold the whole packet. There is no need to divide a packet. This minimizes the overhead problem.

Virtual-cut through is much like the wormhole switching and the main difference is that the header flit can travel ahead of the remaining flits and undergo processing, where the other flits are passing through the network.

2.3.3 NoC Routing:

The NoC routing mechanism is responsible for correct and efficient routing of packets that are traversing the network from source to target [9].

Routing algorithm decides the path that a packet should take to reach to its destination. High performance, load-balance, deadlock-free and livelock-free, fault-tolerant are the desirable properties of a routing algorithm for NoC [8].

Routing scheme can be classified into various categories: The routing can be static or dynamic, distributed or source routing and minimal or non-minimal. In static routing permanent paths from a given source to destination are defined. It may use single path or split the traffic is a predefined way among several paths. In dynamic routing, the traffic between a source target changes its route with time. In distributed routing, each packet carries the destination address [9]. In source routing, the pre-computed routing tables are stored in the NI at the system modules. The power consumption introduced by non-minimal routing prohibitively increases the expensive in the NoC.

2.3.4 Flow Control:

Another research area in NoC is flow control [7]. It determines how network resources, such as channel bandwidth, buffer capacity, and control state, are allocated to a packet traversing the network. The flow control may be classified as buffered or bufferless.

The bufferless flow control has more latency and less throughput than the buffered flow control. The buffered flow control can be further categorized into credit based flow control, ACK/NACK flow control, STALL/GO flow control, T-Error flow control, and Handshaking Signal based flow control.

In credit based flow control, an upstream node keeps count of data transfers, and thus the available free slots are termed as credits. Once the transmitted data packet is either consumed or further transmitted, a credit is sent back. Bolotin et al. used Credit Based flow control in QNOC [7].

In Handshaking signal based flow control, a VALID signal is sent whenever a sender transmits any flit. The receiver acknowledges by asserting a VALID signal after consuming the data flit. Zeferino et al. used handshaking signals in their SoCIN NOC implementation [11].

In the ACK/NACK protocol a copy of a data flit is kept in a buffer until an ACK signal is received. On declaration of ACK, the flit is deleted from the buffer; instead if a NACK signal is asserted then the flit is scheduled for retransmission.

The T-Error flow control scheme is very complex as compared to other flow control mechanisms. It aims at improving the performance at the cost of reliability. Real time systems operating in a noisy environment must avoid the use of this flow control mechanism. None of the present NoC implementations has employed this flow control scheme [7].

2.4 PERFORMANCE ANALYSIS

The implementation and performance of a network on chip can be evaluated by several parameters. They are area consumption, power consumption, bandwidth, throughput and latency [2]. Area and power are related with hardware part of NoC.

The bandwidth refers to the maximum rate of data propagation once a message is in the network. The unit of measures of bandwidth is bit per second (bps) and it usually considers the whole packet, including the bits of the header, payload and tail.

Throughput is defined as the maximum traffic accepted by the network that is the maximum amount of information delivered per time unit.

Latency is the time elapsed between the beginning of the transmission of a message and its complete reception at the target node.

3. NoC SIMULATION TOOLS

Several types of software simulation tools are used and being developed by scientific research teams [14]. Primarily they are used for simulation and synthesis purpose. Few simulation tools addressed in the literatures [14] are given as follows;

1. NS-2 is used for prototyping and simulating conventional computer networks. Since NoCs shares many characteristics with computer networks, NS-2 was widely used by many NoC researchers to simulate NoCs [15-16].
2. Noxim This tool has been proposed by the Computer Architecture team at the University of Catania [17]. It is developed in SystemC language. It allows the user to define a 2D mesh NoC architecture with various parameters including: 1) Network size 2) Buffers size 3) Packet size 4) Routing algorithm 5) Injection rate of packets. Noxim allows the evaluation of NoCs in terms of throughput, latency and power consumption.
3. DARSIM is a NoC simulator which was developed at the Massachusetts Institute of Technology (MIT). This tool allows the simulation of mesh NoC architectures of 2 and 3 dimensions [18].

4. SunFloor - 3D SunFloor: SunFloor is a support tool for NoC design. It can be used at earlier design phases to synthesize the most appropriate topology with these constraints as input (Model, Energy and Space, Design Objectives). From these data, SunFloor generates a system specification ready to be translated into comprehensive architecture, usually in SystemC language and by the intervention of a second tool which is xpipesCompiler. SunFloor 3D is an extension of the later version [19].
5. ORION 2.0 is the successor of the version proposed by a team from Princeton University in 2003 [20]. It is a simulator dedicated primarily to the estimation of power and space for NoCs architectures.

4. CONCLUSION

NoC encompasses a wide spectrum of research, ranging from highly abstract software related issues, across system topology to physical level implementation. This paper gives an overview of state-of-the art network-on-chip. The emerging field of NoC research and design challenges were discussed. Network on Chip is a very active research field with many practical applications in industry. This work focuses on the system, network and link level issues of the communication infrastructure. NoC research areas have to be explored to meet the design challenges in advanced SOC system that becomes leader in microelectronics.

REFERENCES

[1] Erno Salminen, Ari Kulmala and Timo D. Hamalainen, "Survey of Network-on-chip Proposals", White Paper, OCP-IP, 2008.

[2] Erika Cota, Alexandre de Morais Amory and Marcelo Soares Lubaszewski, "*Reliability, Availability and Serviceability of Networks-on-Chip*", Springer, 2012.

[3] Davide Bertozzi, Shashi Kumar and Maurizio Palesi, "Networks-on-Chip: Emerging Research Topics and Novel Ideas", *VLSI Design*, Vol. 2007, Article No. 26454, pp.1-3, 2007.

[4] Tobias Bjerregaard and Shankar Mahadevan, "A Survey of Research and Practices of Network-on-Chip", *ACM Computing Surveys*, Vol. 38, No. 1, 2006.

[5] Erno Salminen, Ari Kulmala and Timo D. Hamalainen, "On Network-on-Chip comparison", *Tenth Euromicro Conference on Digital System Design: Architectures, Methods and Tools*, pp. 503-510, 2007.

[6] C. A. Zeferino, M. E. Kreutz, L. Carro and A. A. Susin, "A study on communication issues for systems-on-chip", *Proceedings of the 15th Symposium on Integrated Circuits and Systems Design*, pp. 121-126,2002.

[7] Ankur Agarwal, Cyril Iskander and Ravi Shankar, "Survey of Network on Chip (NoC) architecture and contributions", *Journal of Engineering Computing and Architecture*, Vol. 3, No. 1, pp. 21-27, 2009.

[8] Yongfeng Xu, Jianyang Zhou and Shunkui Liu, "Research and Analysis of Routing Algorithms for NoC", *Proceedings of 3rd International Conference on Computer Research and Development*, pp. 98-102, 2011.

[9] Giovanni De Micheli and Luca Benini, "*Networks-on-Chips: Technology and Tools*", Morgan Kaufmann Publishers, 2006.

[10] Dan Marconett, "A survey of Architectural Design and Implementation Tradeoffs in Network on Chip Systems", University of California, pp.1-8, 2006.

[11] C. A. Zeferino and A. A. Susin, "SoCIN: A parametric and scalable network-on-chip", *Proceedings of 16th Symposium on Integrated Circuits and Systems Design*, pp. 169-174, 2003.

[12] U. Y. Ogras and R. Marculescu, "Modeling, Analysis and Optimizationof Network-on-Chip Communication Architectures", *Lecture Notes in Electrical Engineering*, Vol. 184, 2013.

[13] O. Tayan, "Network-on-chip: challenges, trends and mechanisms for enhancements", *Proceedings of International Conference on Information and Communication Technologies*, pp. 57-62, 2009.

[14] Ahmed Ben Achballah and Slim Ben Saoud, "A Survey of Network-On-Chip Tools", *International Journal of Advanced Computer Science and Applications*, Vol. 4, No. 9, pp. 61-67, 2013.

[15] M. Ali, Michael Welzl, A. Adnan and F. Nadeem, "Using the NS-2 Network Simulator for Evaluating Network on Chips (NoC)", *International Conference on Emerging Technologies*, pp. 506 – 512, 2006.

[16] NS-2 website. Available: http://nsnam.isi.edu/nsnam/index.php/Main_Page

[17] Noxim website. Available: www.noxim.org

[18] Mieszko Lis, Keun Sup Shim, Myong Hyon Cho, Pengju Ren, Omer Khan and Srinivas Devadas, "DARSIM: A Parallel Cycle-Level NoC Simulator", *6thAnnual Workshop on Modeling, Benchmarking and Simulation*, 2010.

[19] C. Seiculescu, S Murali, L Benini and G. De Micheli, "SunFloor 3D: A Tool for Networks on Chip Topology Synthesis for 3D Systems on Chips", *Proceedings of Design, Automation and Test in Europe Conference & Exhibition*, pp. 9-14, 2009.

[20] Hang-Sheng Wang, Xinping Zhu, Li-Shiuan Peh and S. Malik, "Orion: A Power-Performance Simulator for Interconnection Networks", *Proceedings of 35thAnnual IEEE/ACM International Symposium on Microarchitecture*, pp. 294-305, 2002.

[21] http://en.wikipedia.org/wiki/Network_on_a_chip.