IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015.
www.ijiset.com
ISSN 2348 – 7968

Overview of the technology Network-on-Chip

Mohammad Trik[1], Amir-Masoud Bidgoli[2] and Salam Khazali[3] and Azad Shojaei[4]

[1,3,4]Young Researchers and Elite Club, Sardasht Branch, Islamic Azad University, Sardasht, Iran
[1,3,4]Young Researchers and Elite Club, Urmia Branch, Islamic Azad University, Urmia, Iran
[2]Islamic Azad University Tehran North Branch MIEEE Manchester University Tehran, Iran

Abstract

Network on chip or network on a chip (NoC or NOC) is a communication subsystem on an integrated circuit (commonly called a "chip"),typically between intellectual property (IP) cores in a system on a chip (SoC). NoCs can span synchronous and synchronous clock domains or use uncloked synchronous logic. NoC technology applies networkingtheory 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. This integrated microprocessor has been a landmark in the evolution of computing technology. Whereas it took monstrous efforts to be completed, it appears now as a simple object to us. Indeed, the microprocessor involved the connection of a computational engine to a layered memory system, and this was achieved using busses. Complex application-specific integrated circuits (ASICs) were designed to address-specific applications. These systems required multiprocessing over heterogeneous functional units, thus requiring efficient on-chip communication. On the other side, multiprocessing platforms were developed to address high-performance computation—such as image rendering. Furthermore, the chapter explains variability and design methodologies of NoCs. Dealing with variability is an important matter affecting many aspects of systems-on-chips (SoC) design. The first important issue deals with malfunction containment. Traditionally, malfunctions are avoided by putting stringent rules on physical design and by applying stringent tests on signal integrity before tape out. This approach is conservative in nature and leads to a perfect physical layout of circuits.

Keywords: *NOC technology, System-on-Chip, routing algorithms, performance*

1. Introduction

To meet the growing computation-intensive applications and the needs of low-power, high-performance systems, the number of computing resources in single-chip has enormously increased, because current VLSI technology can support such an extensive integration of transistors. By adding many computing resources such as CPU, DSP, specific IPs, etc to build a system in System-on-Chip, its interconnection between each other becomes another challenging issue. In most System-on-Chip applications, a shared bus interconnection which needs an arbitration logic to serialize several bus access requests, is adopted to communicate with each integrated processing unit because of its low-cost and simple control characteristics. However, such shared bus interconnection has some limitation in its scalability because only one master at a time can utilize the bus which means all the bus accesses should be serialized by the arbitrator. Therefore, in such an environment where the number of bus requesters is large and their required bandwidth for interconnection is more than the current bus, some other interconnection methods should be considered.

Such scalable bandwidth requirement can be satisfied by using on-chip packet-switched micro-network of interconnects, generally known as Network-on-Chip (NoC) architecture. The basic idea came from traditional large-scale multi-processors and distributed computing networks. The scalable and modular nature of NoCs and their support for efficient on-chip communication lead to NoC-based system implementations. Even though the current network technologies are well developed and their supporting features are excellent, their complicated configurations and implementation complexity make it hard to be adopted as an on-chip interconnection methodology. In order to meet typical SoCs or multi-core processing environment, basic module of network interconnection like switching logic, routing algorithm and its packet definition should be light-weighted to result in easily implemental solutions.

2.Background

As the semiconductor processing technology is advanced to sub-nano one, there are several side effects awaited. One of critical issues is a wiring delay. While the speed of basic elements such as gate delay becomes much faster, the wiring delay is growing exponentially as shown in Figure 1 because of the increased capacitance caused by narrow channel width and increased crosstalk. Therefore, if this trend be sustained, the wiring is one of the critical issues to be concerned.

ISSN 2348 – 7968

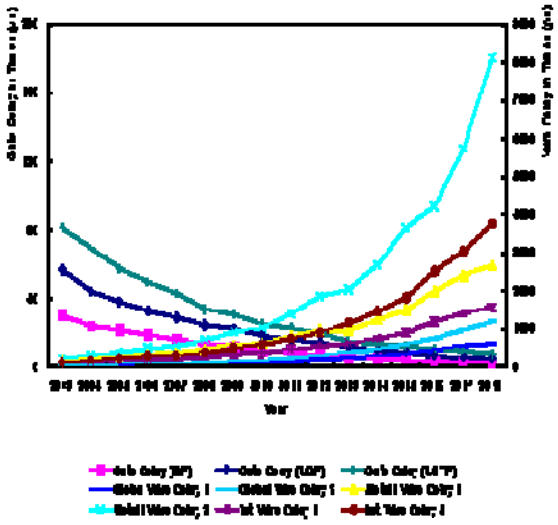

Figure 1. Technical Roadmap in Semiconductor Industry [1]

In communication between several cores in System-on-Chip (SoC) environment, some prevailing mechanisms for this purpose are several bus-based architectures and point-to-point communication methodologies. For simplicity and ease of use, the bus-based architectures are the most common. However, in bus-based architecture, it has fundamentally some limitation in bandwidth, i.e. while the number of components attached to the bus is increased, a physical capacitance on the bus wires grows and as a result its wiring delay grows even further. To overcome the fundamental limitation of scalability in bus-based architectures, some advanced bus architectures such as ARM AMBA [2], OpenCores WISHBONE System-on-Chip (SoC) interconnection [3], and IBM CoreConnect [4], are adopted. The Figure 2 illustrates basic structure of ARM AMBA. As shown in Figure 2, most of advanced bus architectures adopt a hierarchical structure to obtain scalable communication throughput and partition communication domains into several group of communication layers depending on bandwidth requirement such as high-performance, low-performance and so on.

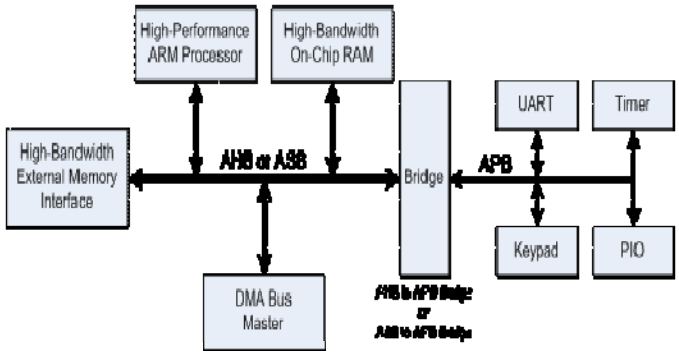

Figure 2. Typical ARM AMBA-bused System

Another approach to exceed such a limitation of communication and overcome such an enormous wiring delay in future technology is to adopt network-like interconnections which is called Network-on-Chip (NoC) architecture. Basic concept of such kind of interconnections is from the modern computer network evolution as mentioned before. By applying network-like communication which inserts some routers in-between each communication object, the required wiring can be shortened. Therefore, the switch-based interconnection mechanism provides a lot of scalability and freedom from the limitation of complex wiring. Replacement of SoC busses by NoCs will follow the same path of data communications when the economics prove that the NoC either reduces SoC manufacturing cost, SoC time to market, SoC time to volume, and SoC design risk or increases SoC performance. According to [5], the NoC approach has a clear advantage over traditional busses and most notably system throughput. And hierarchies of crossbars or multilayered busses have characteristics somewhere in between traditional busses and NoC, however they still fall far short of the NoC with respect to performance and complexity.

The success of the NoC design depends on the research of the interfaces between processing elements of NoC and interconnection fabric. The interconnection of a SoC established procedures has some weak points in those respects of slow bus response time, energy limitation, scalability problem and bandwidth limitation. Bus interconnection composed of a large number of components in a network interface can cause slow interface time though the influence of sharing the bus. In addition the interconnection has a defect that power consumption is high on the score of connecting all objects in the communication. Moreover it is impossible to increase the number of connection of the elements infinitely by reason of the limitation of bandwidth in a bus. As a consequence, the performance of the NoC design relies greatly on the interconnection paradigm.

ISSN 2348 – 7968

Though the network technology in computer network is already well developed, it is almost impossible to apply to a chip-level intercommunication environment without any modification or reduction. For that reason, many researchers are trying to develop appropriate network architectures for on-chip communication. To be eligible for NoC architecture, the basic functionality should be simple and light-weighted because the implemented component of NoC architecture should be small enough to be a basic component constructing a SoC. Even though the basic functionality should be simple, it also satisfies the basic requirement in general communication. On the other hand, to apply the prevailing mobile environment, it should be low-powered. In order to be low powered one has to consider many parameters such as clock rate, operating voltage, and power management scheme.

3.Related works

In designing NoC systems, there are several issues to be concerned with, such as topologies, routing algorithms, performance, latency, complexity and so on. Among these factors, nothing can be independent in deciding an NoC architecture. There are several different kinds of topologies. Guerrier and Greiner [6] have proposed a generic interconnect template called SPIN, where a fat-tree architecture is used to interconnect IP blocks. In this fat-tree, every node has four children and the parent is replicated four times at any level of the tree. Kumar et al. [7] has proposed a mesh-based interconnect architecture called CLICH (Chip-Level Integration of Communicating Heterogeneous Elements). This architecture consists of an m x n mesh of switches interconnecting computational resources (IPs) placed along with the switches. Dally and Towles [8] have proposed a 2D torus as an NoC architecture. The Torus architecture is basically the same as a regular mesh. The only difference is that the switches at the edges are connected to the switches at the opposite edge through wrap-around channels. Karim et al. [9] have proposed the OCTAGON MP-SoC architecture with a basic octagon unit consisting of eight nodes and 12 bi-diretional links. Each node is associated with a processing element and a switch. Communication between any pair of nodes takes at most two hops within the basic octagonal unit. Pande et al. [10] have proposed an interconnect template following a Butterfly Fat-Tree (BFT) architecture where the IPs are placed at the leaves and switches placed at the vertices. As a feasible topology in NoC systems, the mesh is getting popular for its modularity; it can be easily expandable by adding new nodes and links without any modification of the existing node structure. As the mesh nodes can be used as basic components in on-chip communication, they are potentially important components to accomplish a scalable communication model in NoC environment [11]. Another reason behind this popularity is the notion of being partitioned into smaller meshes, which is a desirable feature for parallel applications [12].

Another issue in NoC environment is the routing algorithm. In terms of delivering mechanism, i.e. switching technique, there are different types of switching techniques such as circuit switching, packet switching, and wormhole switching[13]. Switching techniques determine when and how internal switches connect their inputs to outputs and the time at which message components may be transferred along these paths. In circuit switching, a physical path from source to destination is reserved prior to the transmission of the data throughout initialization processes including setup and acknowledgement. The reserved path is held until all the data has been delivered. The advantage of this approach is the reserved network bandwidth for the entire duration of the delivered data. However, with respect to the resource utilization, it ties valuable resources during the transmission of data and the initialization processes cause unnecessary delays. In packet switching, data is divided into fixed-length packets and, instead of establishing a path before sending data in circuit switching, whenever the source has a packet to be sent, it transmits the data. In order to store entire packets in a switch, it requires large-sized buffers. Therefore, the need for storing entire packets in a switch makes the buffer requirement high, resulting in infeasible solution for an SoC environment. In wormhole switching, the packets are further divided into fixed length flow control units or flits, resulting in smaller buffer space requirement in the switches. One drawback of this simple wormhole switching is inability of interleaving or multiplexing distinct messages over a physical channel. However, by applying virtual channels, such channel utilization can be increased. Among several switching techniques, wormhole routing has increasingly been advocated as a method of reducing message routing latency.

In terms of the way of choosing a path among the set of possible paths from source to destination, the routing algorithms are classified as deterministic/oblivious and adaptive ones [14]. The oblivious/deterministic routing algorithms choose a route without considering any information about the networks present condition, resulting in relatively simple design complexity. Adaptive routing algorithms use the state of the network like the status of a node or link, the status of buffers for network resources, or history of channel load information. Adaptive routing algorithms are refined as minimal or fully adaptive routing ones depending on the degree of adaptivity. Even though the adaptive routing algorithms utilize the flexibility in routing paths, the design complexity should be increased. DOR (dimension-ordered routing) [15], ROMM [16], and O1TURN [17] are

ISSN 2348 – 7968

examples of deterministic or oblivious routing algorithms. And some researchers have developed better performance routing algorithms using adaptive routing algorithms [18][19][20][21][22][23].

On the other hand, the adoption of virtual channel (abbreviated to VC) has been prevailing because of its versatility. By adding virtual channels and proper utilization, deadlock-freedom can be easily accomplished. Network throughput can be increased by dividing the buffer storage associated with each network channel into several virtual channels [19], resulting in increase of channel utilization. By proper control of virtual channels, network flow control can be easily implemented[24]. Also to increase the fault tolerance in network, the concept of virtual channel has been utilized [25][26]. However, in order to maximize its utilization, how to allocate virtual channels is a critical issue in designing routing algorithms [27][28].

The network interconnects implement interfaces such as AXI, OCP and DTL to connect IP modules within the NoC. AMBA Extended Interface (AXI) [29] is the next generation, high performance on-chip interface technology developed by ARM to support ARM11 family-class processors. The configurable AXI interconnection components provide data-efficient, highly-optimized link from the processor and data bursting in ARM core-based NoC systems. Furthermore the AXI configurable Interconnect supports a multi-layer topology that guarantees the necessary bandwidth and low latency for all connected IPs and it provides related ARM technologies, such as IEM for voltage and frequency scaling. The Open Core Protocol (OCP) [30] is a plug and play interface for a core having both master and slave interfaces. The OCP signals of the functional IP blocks are packetized by a second interface. All signals are synchronous, simplifying core implementation, integration and timing analysis. It defines a point-to-point interface between two communicating entities and each component acts as the master and the slave. The OCP integrates all inter-core communications, including dataflow and sideband control signals. The Device Transaction Level (DTL) [31] is one of standard for interconnection researched by Philips Semiconductors to interface IPs existing a SoC. The DTL allows easy extension to other future interconnection standard.

Reference

[1] ITRS, *International Technology Roadmap for Semiconductors 2004 Update*, 2004.

[2] AMBA" Specification Rev. 2.0, http://www.arm.com, 1999.

[3] Specification for the: WISHBONE System-on-Chip (SoC) Interconnection Architecture for Portable IP Cores, OpenCore, 2002.

[4] *The CoreConnect Bus Architecture* http://www-03.ibm.com/chips/products/coreconnect/, 1999.

[5] A Comparison of Network-on-Chip and Busses, http://www.arteris.com/noc_whitepaper.pdf, 2005.

[6] P. Guerrier and A. Greiner, "A Generic Architecture for On-Chip Packet-Switched Interconnections," *Proc. Design and Test in Europe (DATE)*, pp. 250-256, Mar. 2000.

[7] S. Kumar et al., "A Network on Chip Architecture and Design Methodology," *Proc. Intel Symp. VLSI (ISVLSI)*, pp. 117-124, 2002.

[8] W. J. Dally and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks," *Proc. Design Automation Conf. (DAC)*, pp. 683-689, 2001.

[9] F. Karim et al., "An interconnect Architecture for Networking Systems on Chips," *IEEE Micro*, vol. 22, no. 5, pp. 36-45, Sept./Oct. 2002.

[10] P. P. Pande et al., "Design of a Switch for Network on Chip Applications," *Proc. Intel Symp. Circuits and Systems (ISCAS)*, vol. 5, pp. 217-220, May 2003.

[11] J. Duato et al., "Interconnection Networks: An Engineering Approach," *IEEE Computer Society Press*, 2003.

[12] H. H. Najaf-abadi et al., "Performance Modeling of Fully Adaptive Wormhole Routing in 2D Mesh-Connected Multiprocessors," *Proc. Intel Symp. Modeling, Analysis, and Simulation of Computer and Telecommunications Systems (MASCOTS)*, pp. 528-534, Oct. 2004.

[13] P. P. Pande et al, "Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures," I*EEE Trans. Computers*, vol. 54, no, 8, pp. 1025-1040, Aug. 2005.

[14] *Principles and Practices of Interconnection Networks*, W. J. Dally and B. Towles, Morgan Kaufmann Publishers, San Francisco, CA, 2004.

[15] H. Sullivan and T. R. Bashkow, "A Large Scale, Homogeneous, Fully Distributed Parallel Machine," *Proc. Symp. Computer Architecture*, pp. 105-117, ACM Press, 1977.

[16] T. Nesson and S. L. Johnsson, "ROMM Routing on Mesh and Torus Networks," *Proc. ACM Symp. Parallel Algorithms and Architectures*, pp. 275-287, ACM Press, 1995.

[17] D. Seo et al., "Near-Optimal Worst-case Throughput Routing for Two-Dimensional Mesh Networks," *Proc. Intel Symp. Computer Architecture (ISCA)*, pp. 432-443, June 2005.

[18] J. Hu and R. Marculescu, "DyAD Smart Routing for Network-on-Chip," *Proc. Design and Automation*, pp. 260-263, ACM Press, 2004.

[19] W. J. Dally et al., "Deadlock-free Message Routing in Multiprocessor Interconnection Networks," *IEEE Trans. Computer*, vol. C-36, no. 5, pp. 547-553, May 1987.

[20] J. Duato, "A New Theory of Deadlock-free Adaptive Routing in Wormhole Networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 4, no. 12, pp. 1320-1331, Dec. 1993.

[21] G. Chiu, "The Odd-Even Turn Model for Adaptive Routing," *IEEE Trans. Parallel and Distributed Systems*, vol. 11, no. 7, pp. 729-738, Jul. 2000.

ISSN 2348 – 7968

[22] C. J. Glass and L. M. Ni, "The Turn Model for Adaptive Routing," *Journal of ACM*, vol. 31, no. 5, pp. 874-902, Sep. 1994.
[23] C. J. Glass and L. M. Ni, "Maximally Fully Adaptive Routing in 2D Meshes," *Proc. Intel Conf. Parallel Processing*, I: 101-104, 1992.
[24] W. J. Dally, "Virtual-Channel Flow Control," *IEEE Trans. Parallel and Distributed Systems*, 3(2): 194-205, Mar. 1992.
[25] R. V. Boppana and S. Chalasani, "Fault-Tolerant Wormhole Routing Algorithms for Mesh Networks," *IEEE Trans. Computers*, vol. 44, no. 7, Jul. 1995.
[26] J. Zhou and F. C. Lau, "Adaptive Fault-Tolerant Wormhole Routing with Two Virtual Channels in 2D Meshes," *Proc. Intel Symp. Parallel Architectures, Algorithms and Networks (ISPAN)*, pp. 142-148, May 2004.
[27] A. S. Vaidya et al., "Impact of Virtual Channels and Adaptive Routing on Application Performance," *IEEE Trans. Parallel and Distributed Systems*, vol. 12, no. 2, pp. 223-237, Feb. 2001.
[28] M. Rezazad and H. Sarbaziazad, "The Effect of Virtual Channel Organization on the Performance of Interconnection Networks," *Proc. Intel Parallel and Distributed Processing Symposium (IPDPS)*, Apr. 2005.
[29] ARM. 2004. AMBA Advanced eXtensible Interface (AXI) Protocol Specifcation, Version 1.0. http://www.arm.com.
[30] OCP International Partnership. Open Core Protocol Specification. 2.0 Release Candidate, 2003.
[31] PHILIPSSEMICONDUCTORS. 2002. Device Transaction Level (DTL) Protocol Specification, Version 2.2.