Design of Optimized Network on-Chip for Reliable Communication

[1]Dr. Jayaprakash. M [2]Manikandan. S [3]Pradeep Kumar. S [4]Sam Jasper. P [5]Prakash. C
[1]Professor [2,3,4,5]Assistant Professor
[1,2,3,4,5]Department of Electrical and Electronics Engineering
[1,2,3,4,5]JCT College of Engineering and Technology, Coimbatore

Abstract

In this paper, a new mesh-typed NoC(Network on Chip) architecture is proposed which aims at enhancing network performance. Networks-on-Chips (NoCs) are a new design paradigm for scalable high throughput communication infrastructures, in Systems-on-Chips (SoCs) with billions of transistors. The idea of NoCs is dividing a chip into several independent clusters connected together by global communication architecture. As the number of cores integrated into System-on-Chip increases, the on-chip communication limits the performance and power consumption in current and next generation SoCs. The resultant NoC uses mesh topology along with virtual channel allocation methodology. The routing algorithm combined with mesh topology improves average latency and saturation traffic load.

Keyword- Systems-on-Chip, Multiprocessor Array, Network-on-Chip (Noc), Mesh Type Noc, Virtual Channel

I. INTRODUCTION

As the trend of device miniaturization continues the number of transistors per chip doubles every couple of years. The increasing density can be used in several ways: the size of the chips can be reduced, individual processing blocks can become more complex thus providing higher processing power, and more functional blocks can be integrated on the same chip. Reducing the size of chips, although beneficial from the cost point of view, cannot be done indefinitely because at a certain point the cost of packaging and terminals would become dominant. The direction that is left and is still promising is the integration functions that are traditionally performed by different devices into a single device.

Integration has several benefits: the cost of several packages is eliminated and the need for connections that would normally go to the outside of the chip is removed. Integration improves performance because communication bandwidth available on chip is significantly higher than off chip. It decreases power consumption as driving external pins uses much more energy than on-chip communication. Another important benefit is the reduced physical size of devices.

Traditionally IP blocks are connected using a single bus or a hierarchy of buses. The parameters of these components could be manually chosen by a skilled engineer and the components themselves could be instantiated from a library to obtain a working system. However, this approach will not scale to designs having tens to hundreds of cores, because companies
Cannot afford increasing the engineering effort per device. Timing constraints become increasingly difficult to meet and verification becomes difficult to perform.

Analyzing the system from the performance point of view also becomes increasingly difficult. While the computation requirements for individual processors can be generally analyzed and verified for many real life applications, the communication performance requirements are less straightforward since the interactions between different IPs need to be taken into account. If the system fails to meet the performance requirements, redesigning the interconnect (or entire SoC) may be a time-consuming and costly operation. It is therefore desirable to have automated tools to dimension and verify the interconnect. These tools start with a high level system or application requirements and automatically generate an interconnect that the system components are attached to. This interconnect may also be verifiable by construction from the correctness and performance points of view.

The remainder of this paper is organized as follows. Section II describes the background of network on chip architecture design and their different topologies. Section III presents the proposed architecture and its design with a switch-by-switch interconnection scheme that can support a backtracking path-setup and a source-synchronous wave pipeline transmission of the data. Section IV. Finally, the conclusion and discussion for further research are given in Section V.

II. BACKGROUND

A. NOC Architecture

The function of an on-chip network is to deliver messages from source node to destination node and there exist many design alternatives to accomplish this job. Depending on the application requirements, how to choose suitable network architecture

remains an open problem in this field of research. Here we discuss the network properties that need to be considered when devising an NoC architecture for specific application needs.

1) Switching Policy

There are two major switching techniques: circuit switching and packet switching. Circuit switching establishes a link between source node and destination node either virtually or physically before a message is being transferred. The link is held until all the data is transmitted. The major advantages of circuit switching are that there is no contention delay during message transmission and its behavior is more predictable, so circuit switching is usually employed when Quality of Service (QoS) is considered.

On the other hand, packet switching transfers messages on a per-hop basis. With packet switching, messages are divided into packets at the source node and then sent into a network. Packets move along a route determined by the routing algorithm and traverse through a series of network nodes and finally arrive at the destination node. Packet switching is utilized in most of NoCs because of its potential for providing simultaneous data communication between many source and destination pairs. Packet switching can be further classified into three classes: store and forward (SAF), virtual cut through (VCT), and wormhole switching. The most popular one for NoC based architectures is wormhole switching because it only requires a buffer size of one flit (flow control unit) so that the area cost of a router can be kept low. In contrast, SAF and VCT require a buffer size of the whole packet which prohibits their adoption.

2) Topology

Topology defines how nodes are placed and connected, affecting the bandwidth and latency of a network. Many different topologies have been proposed, [6], such as mesh, torus, binary tree, Octagon, mixed and custom topology, as shown in Fig. 1

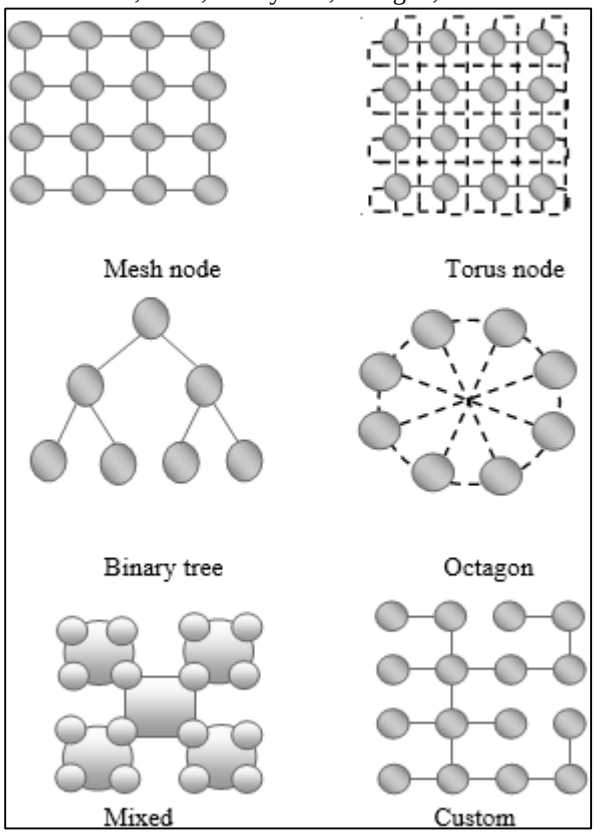

Fig. 1: NoC topologies

Some of the application-specific topology that can offer superior performance while minimizing area and energy consumption [15-17]. The most common topologies are 2D mesh and torus due to their grid-type shapes and regular structure which are the most appropriate for the two dimensional layout on a chip.

3) Routing

Routing is the mechanism responsible for determining the path that a packet traverses from the source node to the destination node. Routing algorithms such as deterministic and adaptive ones have been proposed. With deterministic routing, the path between source-destination pair is fixed, regardless of the current state of the network. On the other hand, an adaptive routing algorithm takes the network state into account when deciding a route, resulting in variation of the routing path with time. For example, it may choose an alternative path if a certain link is congested, therefore, an adaptive routing algorithm has the potential of supporting

more traffic for the same network topology. However, most of the proposed packet-switched NoCs use deterministic routing because of its simplicity and the low area overhead in router design.

B. Multiprocessor Array

The multiprocessor array implements the wormhole packet switching technique and the topology of Multiprocessor array is based on a 2D mesh as shown in Fig. 2. Each node in Multiprocessor array consists of a router and a local IP which can be a CPU, DSP, memory block, or application-specific logic. The router connects with its four neighboring routers via six bidirectional links.

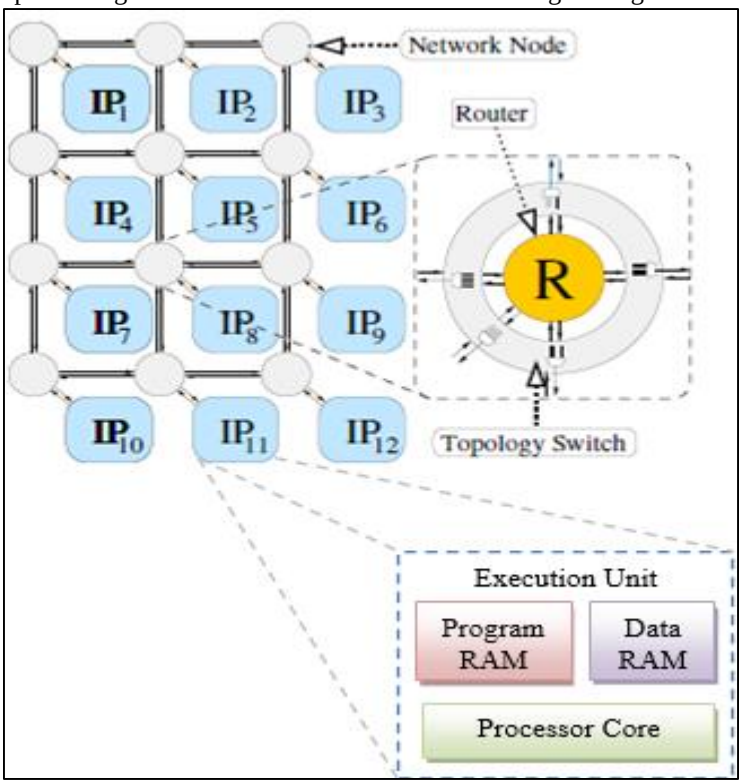

Fig. 2: A 4x4 multiprocessor network and its node composition

A key feature of the Multiprocessor array architecture is the use of two separate vertical links which are employed to construct a deadlock-free network [19]. The Multiprocessor array network is actually composed of two disjoint sub-networks. One sub-network is responsible for delivering east-bounded packets while the other one is for west-bounded packets. Therefore, cycles in the resource dependence graph [14] and prevent deadlocks from happening. This design technique reduces the design complexity of the router because there is no need for a deadlock aware routing algorithm. To increase network performance, Multiprocessor array utilizes an adaptive XY routing algorithm. When an output port is congested, or the output buffer is full, the router selects an alternative output port for packets. Therefore, the link utilization is balanced and network performance improves.

III. PROPOSED ON-CHIP NETWORK TOPOLOGY

A. Topology

In practice, mesh topology is widely used due to its regularity and ease of layout in conventional 2-D chips [1-3]. Torus topology has half the network diameter and twice the number of bisection connections compared to a mesh when accommodating the same number of PEs. For this reason, we consider folded-torus, the laid-out version of torus on 2-D chip, as the topology of interest. We propose to use a dual-lane torus as the topology of our NOC. The "dual-lane" is adopted as a tradeoff between the area overhead and the path diversity of the proposed NOC. [1-3]

B. Switching model

Networks on chip can be split in two large categories: circuit-switching and packet-switching networks. Circuit switching networks allocate relatively long-lived connections between source and destination, and provide high bandwidth and low latency over these connections. In contrast, packet switching networks use arbitration at each network node and for each data packet. Under low network load, packet switching networks provide good latency. One concern for packet switching networks is the "saturation point," where an increase in traffic causes a disproportionate increase in latency.

C. Proposed Network on Chip

In Network on chip is typically less constrained than in bus based interconnects. In particular, unlike a bus hierarchy, a network on chip is not required to be a tree topology. The NoC topology can be adapted to the chip floor plan and can be optimized to take into account the communication requirements between IPs.

IV. DESIGN AND IMPLEMENTATION

The proposed on chip network which uses mesh topology with round robin approach and virtual channel allocation scheme.
The proposed routing protocol constituting shows guaranteed throughput which supports synchronized communication between the processors. Fig. 3(a) shows the output waveform of proposed probing mesh architecture, and Fig. 3(b) shows the on chip network area utilization.

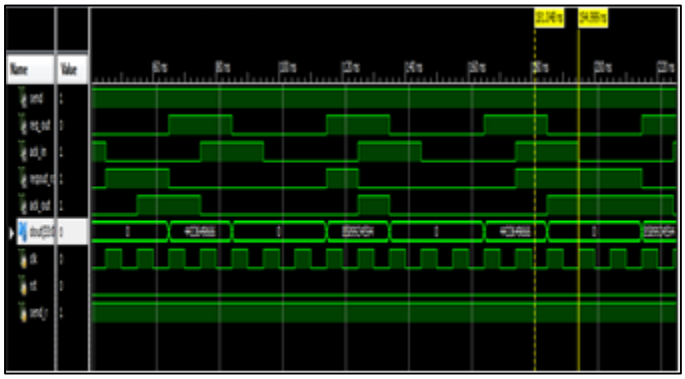

Fig. 3 (a): Output Waveform for proposed NoC

Device Utilization Summary (estimated values)			
Logic Utilization	Used	Available	Utilization
Number of Slices	4287	4656	92%
Number of Slice Flip Flops	4252	9312	45%
Number of 4 input LUTs	6975	9312	74%
Number of bonded IOBs	190	158	120%
Number of GCLKs	1	24	4%

Fig. 3 (b): Synthesis Result for proposed NoC

V. CONCLUSION

This paper presented a FPGA implementation of on chip network for a mesh type NoC. The mesh-typed NoC approach totally reduces the latency through an improved topology. The proposed router architecture is simple to implement yet can achieve the required packet collision avoidance. The proposed routers have significantly less area than other NoC routers. In addition, the simulation results shows that it is more area efficient way for increasing the network performance than using large buffers.

REFERENCES

[1] Abbas Eslami Kiasari, Zhonghai Lu, Axel Jantsch, “An Analytical Latency Model for Networks-on-Chip”, IEEE Transactions on Very Large Scale Integration Systems, Vol. 21, No. 1, 2013.
[2] Phi-Hung Pham, Phuong Mau, Jungmoon Kim, and Chulwoo Kim, “An On-Chip Network Fabric Supporting Coarse-Grained Processor Array”, IEEE Transactions on Very Large Scale Integration Systems, Vol. 21, No. 1, 2013.
[3] Phi-Hung Pham, Phuong Mau, Jungmoon Kim, and Chulwoo Kim, “Design and Implementation of an On-Chip Permutation Network for Multiprocessor System-On-Chip”, IEEE Transactions on Very Large Scale Integration Systems, Vol. 21, No. 1, 2013.
[4] Phi-Hung Pham, Jongsun Park, Member, Phuong Mau, Chulwoo Kim, “Design and Implementation of Backtracking Wave-Pipeline Switch to Support Guaranteed Throughput in Network-on-Chip”, IEEE Transactions on Very Large Scale Integration Systems, Vol. 20, No. 2, 2012.
[5] S. Vangal et al., "An 80-Tile 1.28TFLOPS Network-on-Chip in 65nm CMOS," Solid-State Circuits Conference, 2007. Digest of Technical Papers. IEEE International, pp. 98-589, 2007.
[6] W. J. Dally and B. Towles, "Route packets, not wires: on-chip interconnection networks," Design Automation Conference, 2001. Proceedings, pp. 684-689, 2001.

[7] L. Benini and G. De Micheli, "Networks on chip: a new paradigm for systems on chip design," Design, Automation and Test in Europe Conference and Exhibition, 2002. Proceedings, pp. 418-419, 2002.
[8] T. Bjerregaard and S. Mahadevan, "A survey of research and practices of Network-on-chip," ACM Computing Surveys, vol. 38, pp. 1, 2006.
[9] E. Salminen et al., "Survey of Network-on-Chip proposals," White Paper, OCP-IP, March 2008.
J. H. Bahn, S. E. Lee, Y. S. Yang, J. Yang and N. Bagherzadeh, "On Design and Application Mapping of a Network-on-Chip(NoC) Architecture," Parallel Processing Letters, vol. 18, pp. 239-255, 2008.
[10] G. Varatkar and R. Marculescu, "Traffic analysis for on-chip networks design of multimedia applications," in DAC '02: Proceedings of the 39th Conference on Design Automation, 2002, pp. 795-800.
[11] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," Networking, IEEE/ACM Transactions on, vol. 2, pp. 1-15, 1994.
[12] M. S. Taqqu, W. Willinger and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," SIGCOMM Computer Communication Rev., vol. 27, pp. 5-23, 1997.
[13] D. R. Avresky, "Performance evaluation of the ServerNet(R) SAN under self-similar traffic," Parallel and Distributed Processing, 1999. 13th International and 10th Symposium on Parallel and Distributed Processing, 1999. 1999 IPPS/SPDP. Proceedings, pp. 143-147, 1999.
[14] W. J. Dally, Principles and Practices of Interconnection Networks. Morgan Kaufmann, 2004.
[15] K. Chang, J. Shen and T. Chen, "Evaluation and design trade-offs between circuit-switched and packet-switched NOCs for application specific SOCs," in DAC '06: Proceedings of the 43rd Annual Conference on Design Automation, 2006, pp. 143-148.
[16] P. Marchal, D. Verkest, A. Shickova, F. Catthoor, F. Robert and A. Leroy, "Spatial division multiplexing: a novel approach for guaranteed throughput on NoCs," Hardware/Software Codesign and System.

