Evaluation of MP-SoC Interconnect Architectures: a Case Study

Partha Pratim Pande, Cristian Grecu, Michael Jones, André Ivanov, and Res Saleh
SOC Research Lab
Department of Electrical and Computer Engineering
University of British Columbia
2356 Main Mall Vancouver, BC, V6T 1Z4 Canada
Email: {parthap, grecuc, michaelj, ivanov, res}@ece.ubc.ca

Abstract

Multi-Processor (MP-SoC) platforms are emerging as the latest trend in SoC design. These MP-SoCs consist of a large number of IP blocks in the form of functionally heterogeneous embedded processors. In this new design paradigm, IP blocks need to be integrated using a structured interconnect template, for example, according to high-performance parallel computing architectures. A formal evaluation process is required before adopting a specific parallel architecture to SoC domain. Here, we propose an evaluation methodology based on performance metrics that include latency, throughput and silicon area requirements. As a case study, we present the results of such an evaluation for two MP-SoC interconnect topologies, i.e., the MESH and the Butterfly Fat-Tree (BFT). This evaluation methodology can be extended to any other SoC interconnect topology without loss of generality.

Keywords*: MP-SoC, Mesh, BFT, Latency, Throughput.*

1. Introduction

According to ITRS [1], the realization of complex Systems on a Chip (SoCs) consisting of billions of transistors fabricated in technologies characterized by 65 nm feature size and less will soon be reality. The emergence of SoC platforms consisting of large, heterogeneous sets of embedded processors is imminent. A key element of such multiprocessor SoC (MP-SoC) platforms [2] is the *interconnect topology*. We specifically propose that the on-chip interconnect topology should resemble the interconnect architecture of high-performance parallel computing systems. The common characteristic of these kinds of architectures is that the functional IP blocks communicate with each other through the help of intelligent switches. As such, the switches can be considered as *infrastructure* IPs (I^2Ps) [3] providing a robust data transfer medium for the functional IP modules.

There are different possible interconnect architectures well documented in the parallel processing domain. However, only a few of them would be suitable for SoC implementation. One of these architectures is the Butterfly Fat Tree (BFT) [4]. Another possible interconnect architecture for SoC design is the MESH as proposed in [5] [6].

In this paper, our main contribution is the establishment of a performance/cost evaluation methodology for MP-SoC interconnects. As a case study this methodology is applied and experimental results are shown for the two different MP-SoC architectures, mentioned above.

2. Related Work

A few on-chip micro-network proposals for SoC integration can be found in the literature. Sonic's Silicon Backplane [7] is one example. Kumar [5] and Dally [6] have proposed mesh-based interconnect architectures. In this case, the number of switches is equal to the number of functional IPs. Guerrier and Greiner [8] proposed the use of a tree- based interconnect (SPIN) and addressed system level design issues.

In [4] and [9], we have described an interconnect architecture for a networked SoC based on a BFT architecture, as well as the associated design of required switches and addressing mechanisms.

The precise focus of this paper is to present and discuss a formal evaluation methodology allowing for objective comparisons of different architectures. Such an analysis can prove invaluable for SoC integrators faced with having to select a suitable MP-SoC interconnect template.

3. Performance Metrics

In a communication-centric design methodology, data is transferred among the IP blocks in the form of packets. The need for storing entire packets in a switch makes the buffer requirement high in the case of conventional packet switching. In wormhole switching, the packets are divided into fixed-length flow control units (flits) and the input and output buffers should be able to store only a few flits. As a result, the buffer space requirement in the switches can be small compared to that generally required for conventional packet switching. One drawback of this simple wormhole switching method is that the transmission of distinct messages cannot be interleaved or multiplexed over a physical channel. This will decrease channel utilization if a flit from a given packet is blocked in a buffer. By introducing virtual channels in the input and output ports, channel utility can be increased considerably [9]. If a flit belonging to a particular packet is blocked in one of the virtual channels, then flits of alternate packets can use the other virtual channel buffers, and hence, ultimately, the physical channel.

It is desirable that a MP-SoC interconnect architecture exhibits low latency and high throughput. The additional area overhead due to the infrastructure IPs should be reasonably small.

3.1 Latency

Latency is defined as the time (in clock cycles) that elapses from between the occurrence of a message header injection into the network at the source node and the occurrence of a tail flit reception at the destination node [10].

In order to reach the destination node, the flits have to travel through a path consisting of a set of *stages*. Depending on the source/destination pair and the routing algorithm, each message may have a different latency. We use the average latency as a performance metric in our evaluation methodology.

Let P be the total number of messages reaching their destination IPs, and let L_i be the latency of each message, where i

Proceedings of the 4th IEEE International Workshop on System-on-Chip for Real-Time Applications (IWSOC'04)
0-7695-2182-7/04 $ 20.00 IEEE

varies from 1 to *P*. The average latency, *Lat,* is then calculated according to the following:

$$Lat = \frac{\sum_{1}^{P} L_i}{P}$$

3.2 Throughput

We define *throughput* [TP] as follows:

$$TP = \frac{(Total\ messages\ completed) \times (Message\ length)}{(Number\ of\ IP\ blocks) \times (Total\ time)}$$

Total messages completed refers to the number of whole messages that successfully arrive at their destination IPs, *Message length* is measured in flits, *Number of IP blocks* is the number of functional IP blocks, and *Total time* is the time (in clock cycles) that elapses between the occurrence of the first message generation and the last message reception. Thus, throughput is measured as the fraction of the maximum load the network is capable of physically handling. A throughput *TP* = 1 corresponds to all end nodes receiving one flit every cycle. Realistically, *TP* < 1 since it is improbable that all possible destinations be active at each cycle. Accordingly, throughput is measured in flits/second.

3.4 Area Requirements

To evaluate the feasibility of these interconnect schemes we need to study their silicon area requirements. As the switches form an integral part of the active components of these infrastructures, it is important to determine the amount of relative silicon area they consume. The switches have two main components: the storage buffer, and logic to implement routing and flow control. The storage buffers are the FIFOs at the inputs and outputs of the switch. Another source of silicon area overhead is the area for the inter-switch wires, which, depending on their lengths may have to be buffered through repeater insertion to keep the inter-switch delay within one clock cycle [11]. Consequently, this additional buffer area has to be accounted for.

4. Evaluation Methodology

The communication-centric parameters that need to be considered when evaluating a MP-SoC interconnect are latency and throughput. The area overhead due to the I^2Ps should be considered as well since their sole purpose is to transfer data among functional IP blocks.

We developed a flit-level event-driven wormhole routing simulator to study the characteristics of the communication-centric parameters of the interconnect infrastructure.

To estimate the silicon area consumed by the switches, we developed their VHDL models and synthesized them using a fully static, standard cell-based, CMOS 0.13 μm technology.

5. Interconnect Architectures

In [4], we first proposed a novel interconnect template to integrate numerous IPs of an SoC following a butterfly fat-tree architecture. In our network, the IPs are placed at the leaves and switches are placed at the vertices. Figure 1 illustrates the physical placement of a butterfly fat tree with ***64*** IPs. In the limit, the number of switches converges to *N/2* for a system size of *N,* as *N* grows arbitrarily large. In the case of 64 IPs the number of switches is 28 as shown in Figure 1.

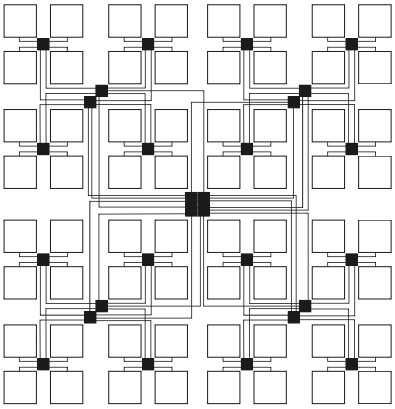

Figure 1: Floorplan of a 64-IP-BFT network.

In a mesh topology, IP blocks are arranged in a two dimensional array structure where each switch (infrastructure IP) is connected to four neighboring switches and a functional IP module. Figure 2 shows the physical placement of 64 functional IP blocks in a mesh based on [5] [6]. In both Figs.1 and 2, the darker blocks denote the switches, and the white squares represent the functional IPs. It is clear that the number of switches in a mesh is equal to the number of interconnected functional IP blocks.

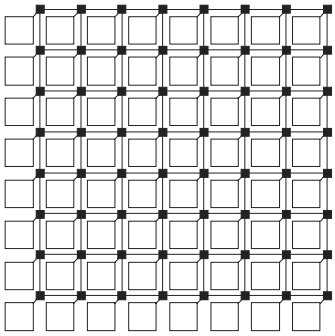

Figure 2: Floorplan of a 64-IP-MESH network.

7. Case Study

We applied our evaluation methodology to two different MP-SoC interconnect platforms, e.g. MESH [6] and BFT [4].

Data were obtained by using a 2500 lines event-driven wormhole routing simulator written in C++.

Uniform, localized, and bit-reversal traffic patterns can be chosen as well. When injections occur, a source IP is selected randomly, according to a uniform distribution. The mean time between packet injections can be specified to alter the overall load on the network. The destination IP selection depends on the traffic pattern adopted. All packets are assumed to have a constant length.

Switch parameters can also be specified. These include input/output port buffer depths (in flits), number of ports, and the number of virtual channels per switch port.

All resource contention is handled without bias such that granting of resources to packets is done on a first come, first-serve basis. The routing algorithms employed are LCA (Least Common Ancestor) determination for the BFT and X-Y routing for the MESH [10].

7.1 Communication-centric Parameters

Latency: In Figure 3 we have plotted the average latency, varying the injection load and the number of virtual channels. Figure 3 indicates a better performance of the BFT in terms of latency till the network saturates.

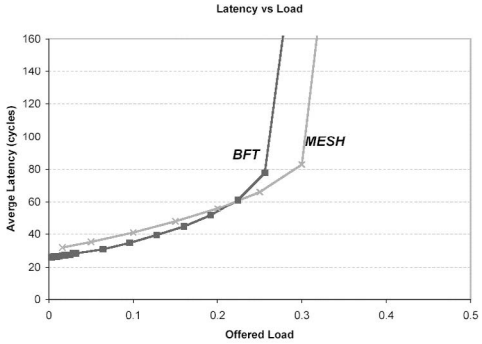

Figure 3 (a): Latency comparison (varying injection load)

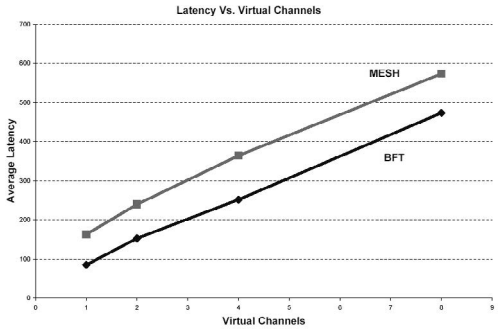

Figure 3 (b): Latency comparison (varying virtual channels)

Throughput: The throughput of the communication infrastructure generally depends on the traffic pattern. Measuring throughput under random traffic with uniform distribution assumptions is an accepted metric [10] for evaluating parallel systems. Figure 4(a) shows the variation of throughput with the number of virtual channels for both the BFT and MESH topologies, under the stated distribution assumptions.

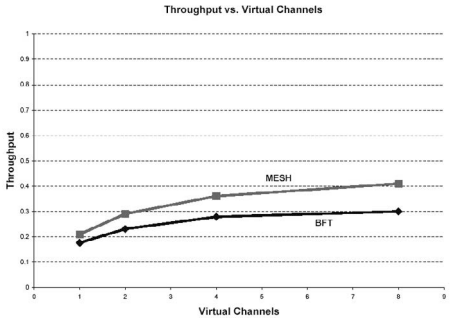

Figure 4(a): Throughput comparison under uniform traffic.

It turns out however, that this uniform traffic assumption is not very realistic in an SoC environment, since different functions will be mapped to different parts of the SoC and they will exhibit highly localized patterns. Hence, we studied the effect of traffic localization on throughput. We considered a very simple case of localization where local messages travel from source to the set of the nearest destinations. In the case of BFT localized traffic is constrained within a single sub-tree while in the case of MESH it is constrained within the four destinations placed at the shortest Manhattan distance [10]. We define the fraction of localization as the ratio between the local traffic and the total traffic. Fig 4(b) shows the effect of traffic localization on throughput for both topologies.

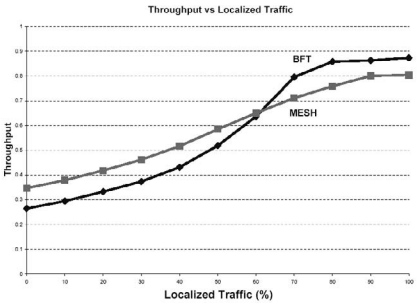

Figure 4(b): Throughput comparison under localized traffic.

From Figure 4(b), we conclude that though the BFT has a lower throughput under uniform traffic, in the presence of localized traffic, its throughput is higher if the fraction of localization is beyond 60%.

7.2 Area Requirements

From our initial implementation, we deduce that within a switch the buffer area significantly dominates over the logic [9]. The buffer area, in turn, largely depends on the number of virtual channels and the flit length. The proposed packets consist of a header flit, one or more data flits, and a tail flit. The header, data, and tail flit structures are as shown in Figure 5.

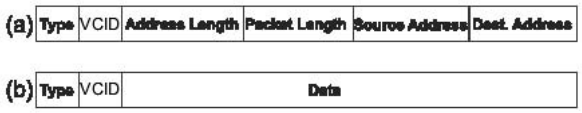

Figure 5: Packet structure (a) Header flit; (b) Data and Tail flit.

The first field denotes the flit type, namely *header, data* or *tail*. The second field contains the virtual channel identifier (VCID). The third field denotes the address length, which is dependent on the number of SoC IP blocks. The fourth field contains packet length information, i.e., the number of flits in the corresponding packet. The next two fields give source and destination addresses.

From Figure 4(a), if the number of virtual channels is increased beyond four, then there is a trend towards throughput saturation. A system with four virtual channels strikes an appropriate balance between high throughput and conservation of silicon area.

The number of IPs in a single SoC varies from one technology node to another, and consequently, the number of bits required to address the IPs will also vary. We consider that each functional IP block consists of 100K gates [11]. In a networked SoC, IPs can be divided into two groups, *functional* and *infrastructure* IPs (switches). Through RTL level design and synthesis, we found that the switches consist of approximately 30K

gates. Consequently, the distribution of functional and I²Ps depends on their respective sizes and interconnect topology. Letting A_{FIP} denote the area of the 100K gates functional IP blocks and A_{I^2P} denote the area of the switches then

$$Area_{chip} = N_1 \cdot A_{FIP} + N_2 \cdot A_{I^2P}$$

where N_1 and N_2 are the number of functional and infrastructure IPs, respectively, and $Area_{chip}$ is the total area of the SoC under consideration. For a BFT, $N_1 = 2N_2$, and for a MESH, $N_1=N_2$. These help us in determining the distribution of functional and infrastructure IP blocks in a SoC.

Functional IPs govern the number of bits required to denote each of address length, source address and destination address fields. We have considered each packet to be consisting of 16 flits [10]. Two bits are needed to specify each of *Flit Type* and *VCID,* and four bits will be sufficient to denote the packet length in each technology node, irrespective of the topologies. The number of bits required to denote the flit type, VCID, and the packet length are topology- and technology-independent. As the number of functional IP blocks varies with interconnect topologies and technology nodes, the size of the address length, source and destination address fields will vary accordingly. Table 1 shows the number of bits in the header flit in the BFT and MESH architectures. The lengths of the body flits (data or tail) are kept equal to the header flit length.

Table 1: Flit lengths in different technology nodes.

Tech. node	Address Length		Source Address		Destination Address		Header Flit Length	
	BFT	MESH	BFT	MESH	BFT	MESH	BFT	MESH
130 nm	4	4	9	9	9	9	30	30
90 nm	4	4	10	10	10	10	32	32
65 nm	4	4	12	11	12	11	36	34
45 nm	4	4	13	13	13	13	38	38
32 nm	4	4	14	13	14	13	40	38

The other contributing factors to the area overhead are the inter-switch repeaters. The wire length between switches in the BFT architecture depends on the levels of the switches. Consequently to keep the inter-switch wire delay within one clock cycle some of them need to be buffered [12]. In a MESH, all the inter-switch wires are of equal length and their delay is always within one clock cycle [12]. Consequently, no repeater insertion is required.

The inter-switch channel width is assumed to be equal to the size of the flit, specified in the last column of Table 1.

The silicon area overhead in different technology nodes can be estimated for both the BFT and MESH interconnect architectures, as the sum of the area due to the switches ($Area_{I^2P}$) and repeaters ($Area_{repeaters}$).

$$Area_{overhead} = Area_{I^2P} + Area_{repeaters}$$

Figure 6 indicates the silicon area overhead across different technology nodes for both platforms.

In terms of silicon requirements, both platforms require a reasonably low area (from 9% in 130 nm node to 14.2% in 32 nm node); however the BFT interconnect architecture offers a more efficient alternative than MESH (20% less area on the average).

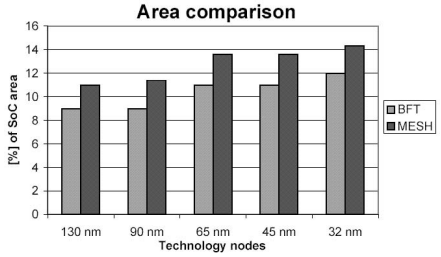

Figure 6: Infrastructure IP area overhead comparison.

8. Conclusions

In this paper, we establish a baseline procedure for performance/cost evaluation of the interconnect architectures to be adopted by the evolving MP-SoC platforms. The interconnect topologies need to be assessed in terms of two types of metrics, communication-centric parameters and the silicon area overhead due to the infrastructures. We presented a case study that illustrates the evaluation of two different topologies, namely the BFT and the MESH.

9. Acknowledgments

The authors wish to thank Micronet, Gennum and NSERC for their financial support and the CMC for providing access to CAD tools.

10. References

[1] ITRS 2003 Documents
http://public.itrs.net/Files/2003ITRS/Home2003.htm

[2] P. Magarshack, P.G. Paulin, "System-on-Chip Beyond the Nanometer Wall", *Proceedings of DAC'03,* June 2-6,2003, Anaheim, USA.

[3] Y. Zorian, "Guest editor's introduction: what is infrastructure IP?" *IEEE Design & Test of Computers*, Volume: 19 Issue: 3 , May-June 2002 pp. 3 -5.

[4] P. P. Pande, C. Grecu, A. Ivanov, R. Saleh, "Design of a Switch for Network on Chip Applications", Proceedings *of ISCAS*, Bangkok, May 2003 Vol. V, pp. 217-220.

[5] S. Kumar, et al, " A Network on Chip Architecture and Design Methodology," *Proceedings of ISVLSI,* pp. 117-124, 2002.

[6] W. J. Dally, and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks", *Proceedings of DAC 2001*, pp. 683-689, Las Vegas, Nevada, USA, June 18-22, 2001.

[7] D. Wingard, " MicroNetwork-Based Integration for SoCs", *Proc. DAC 2001,* pp. 673-677, Las Vegas, Nevada, USA, June 18-22, 2001.

[8] P. Guerrier, A. Greiner**,** "A generic architecture for on-chip packet-switched interconnections", *Proceedings of Design, Automation and Test in Europe Conference and Exhibition 2000*, pp. 250 –256.

[9] P. P. Pande, C. Grecu, A. Ivanov, R. Saleh, "High-Throughput Switch-Based Interconnect for Future SoCs", *Proceedings of 3rd IEEE International Workshop on SoC for Real Time Applications*, 2003, pp. 304-310, Calgary, Canada.

[10] J. Duato, S. Yalamanchili, L. Ni, *Interconnection Networks – An Engineering Approach*, Morgan Kaufmann, 2002.

[11] D. Sylvester, K. Keutzer, "Impact of Small Process Geometries on Microarchitectures in Systems on a Chip", *Proceedings of the IEEE,* Vol. 89, No. 4, April 2001, pp. 467-489.

[12] Cristian Grecu, Partha Pratim Pande, Andre Ivanov, Res Saleh, "A Scalable Communication-Centric SoC Interconnect Architecture," IEEE International Symposium on Quality Electronic Design, ISQED 2004 San Jose, California, USA, 22-24 March, 2004.