ogies, 16 IP and 64 IP cores

latency (μs)		
8X8 Torus	Binary tree	Butterfly Fat Tree
391.135	2401.728	1603.329
391.135	2403.249	1604.348
391.135	2408.624	1608.223
391.135	2411.325	1609.11

al latency for 64 IP nodes

Roadmap for Semiconductors,

Globally Asynchronous Locally
e, 1999, USA.
p Submicron", Proc. of the Int.

Submicron II: A global wiring
pp.193-200.
nted Software, Reading, MA,

lesign methodology", Proceedings

ces of Interconnection Networks,

nect Architecture, Proc. of ISQED

/chips/products/coreconnect/
AC 2001, USA, 2001, pp.673-677.
n-Chip Designs, Proc. of the First
pp.116-120, 2002.
t-switched interconnections, Proc.
tion 2000, pp. 250-256.
etworking systems on chips, IEEE

SCAS, pp.217-220, May, 2003.
r Future SoCs, Proc.3rd IEEE Int.
ada, pp.304-310, June-July, 2003.

Научни трудове на Съюза на учените в България – Пловдив Серия В. Техника и технологии, том XIII., Съюз на учените, сесия 5 - 6 ноември 2015 Scientific Works of the Union of Scientists in Bulgaria-Plovdiv, series C. Technics and Technologies, Vol. XIII., Union of Scientists, ISSN 1311-9419, Session 5 - 6 November 2015.

МЕТОДИ ЗА ТЕСТВАНЕ НА ПРОИЗВОДИТЕЛНОСТТА НА NOC-БАЗИРАНИ ЕТЕРНЕТ СМАРТ СУИЧОВЕ

Иле Димитриевски, Валентин С. Моллов
Технически Университет – София, катедра „Компютърни системи"

PERFORMANCE TESTING METHODS FOR NOC-BASED SMART ETHERNET SWITCHES

Ile Dimitrievski, Valentin S. Mollov
Department of Computer Systems, Technical University of Sofia, Bulgaria

Abstract: Nowadays, when Networks on Chip (NoC) based single-chip networking devices are designed so, to achieve their maximal performance methods appropriate methods for testing of these methods must to be developed.Performance of the NoC-based Ethernet smart switcheshas been rapidly improved andthey are required to fulfill some requirements likelowest possible time delay andoverall latency,an increased traffic speed throughthe network switch, and also an increasedbandwidth and throughput. The state-of-the-art methods for fabric testing of the performance onNoCbased smart Ethernet switchesarepresented. Performance of the differentalgorithms for switching in NoC based smart Ethernet switching will be presented anddiscussed.An overview of selected methodswill be performed and an introduction into simulating of these performance methods will be given.

Keywords: NOC, average latency, message throughput, energy

1.Introduction

The NoC's design methodology is expected to be revolutionary changed during the next years.According to related reference papers [4,5,6], the NoC's platforms in future will consist of large set of embedded processors. On these NoC's numerous IP cores will be integrated performing various functions and working on different clock frequencies. Basic NoC structure is given on Fig.1.One of the main problems associated with the future NoC's design occurs from the non scalability of global wires and delay caused by these lines. Global wires that carry signals across the chip and their length, does not scale with the technology scale.For a relatively long bus line, the intrinsic and parasitic resistance and capacitance can be quite high.

2. Related works

The most frequently used on-chip interconnect architecture is the shared medium arbitrated bus, where all communication devices share the same transmission medium. The advantages of the

shared-bus architecture are simple topology, low area cost, and extensibility. In this paper the basic topologies of NoC's will be presented and they are given in Fig.1.

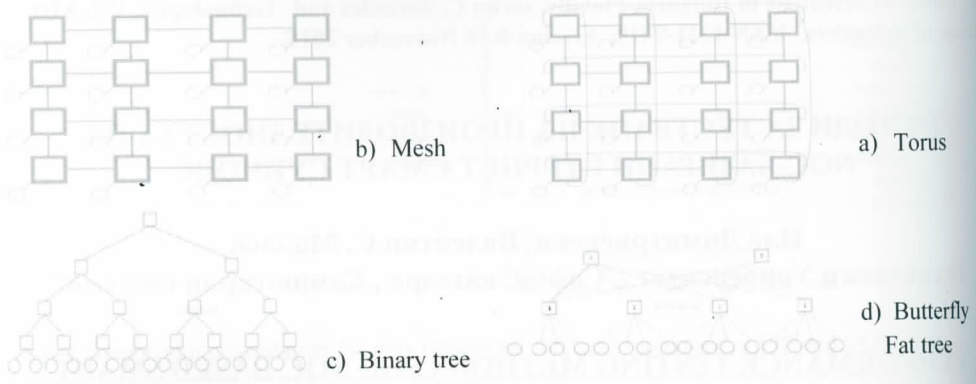


Fig.1Basic NoC topologies

3. Performance metrics

To compare and contrast different NoC architectures, a standard set of performance metrics can be applied [22], [27],[1]. For example, the NoC interconnectarchitecture exhibits high throughput, low latency, energy efficiency, and low area overhead. In today's power constrained environments, it is critical to be able to identify the most energy efficient architectures and to be able to quantify the energy-performance trade-offs [1]. Generally, the additional area overhead due to the infrastructure IPs should be reasonably small. We now describe these metrics in more detail.

3.1 Message Throughput

The performance of a digital communication network is characterized by its bandwidth in and the measurement unit is bits/sec. However, in this case we are more concerned here on the rate that the message traffic can be sent across the network and, so, *throughput* is a more appropriate metric. Throughput can be defined in a different ways depending on the specifics of the implementation, i.e. topologies of the NoC. In general, for message passing systems, definition about message throughput, *TP*, it can be given:

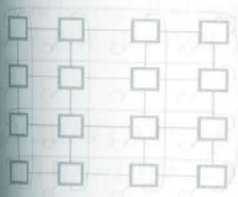
$$TP = \frac{(Total\ messages\ complited)\ x\ (Message\ length)}{(Number\ of\ IP\ blocks)\ x\ (Total\ time)} \quad (1)$$

where *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 involved in the communication, and *Total time* is the time (measured in clock cycles) that elapses between the occurrence of the first message generation and the last message reception. Thus, the message throughput is measured as the fraction of the maximum load that the network is capable of physically handling. An overall throughput of TP=1 corresponds to all end nodes receiving one flit every cycle. Accordingly, throughput is measured in flits/cycle/IP. Throughput signifies the maximum value of the accepted traffic and it is related to the peak data rate sustainable by the system[1].

3.2 Transport Latency

Latency is defined as the time (in clock cycles) that elapses 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 [7]. We refer to this simply as latency in the remainder of this paper. In order to reach the destination node from some starting source node, flits must travel through a path

ea cost, and extensibility. In this paper the basic given in Fig.1.



a) Torus

d) Butterfly Fat tree

C topologies

ics, a standard set of performance metrics can be terconnectarchitecture exhibits high throughput, ea overhead. In today's power constrained he most energy efficient architectures and to be [1]. Generally, the additional area overhead due . We now describe these metrics in more detail.

rk is characterized by its bandwidth in and the we are more concerned here on the rate that the , so, *throughput* is a more appropriate metric. pending on the specifics of the implementation, ge passing systems, definition about message

$$\frac{\text{ed) x (Message length)}}{\text{cks) x (Total time)}} \quad (1)$$

er of whole messages that successfully arrive n flits, *Number of IP blocks* is the number of and *Total time* is the time (measured in clock rst message generation and the last message as the fraction of the maximum load that the ll throughput of TP=1 corresponds to all end throughput is measured in flits/cycle/IP. pted traffic and it is related to the peak data

apses between the occurrence of a message nd the occurrence of a tail flit reception at the in the remainder of this paper. In order to ce node, flits must travel through a path

consisting of a set of switches and interconnect, called *stages*. Depending on the source/destination pair and the routing algorithm, each message may have a different latency. There is also some overhead in the source and destination that also contributes to the overall latency. Therefore, for a given message i, the latency L_i is:

$$L_i = sender\ overhead + transport\ latency + receiver\ overhead. \quad (2)$$

We use the average latency as a performance metric in our evaluation methodology. The average latency is crucial for evaluating of the performance of NoC.P will be the total number of messages reaching their destination IPs and,L_i is the latency of each message i, where i ranges from 1 to P. The average latency, L_{avg}, is then calculated according to the following:

$$L_{avg} = \frac{\sum_1^P L_i}{P}. \quad (3)$$

3.3 Energy

When flits travel on the interconnection network, both the interswitch wires and the logic gates in the switches toggle and this will result in energy dissipation and this definition was givenin reference [1]. In this paper, we are concerned with the dynamic energy dissipation caused by the communication process in the network. The flits from the source nodes need to traverse multiple hops consisting of switches and wires to reach destinations. We are determine the energy dissipated by the flits in each interconnect and switch hop. The energy per flit per hop is given by:

$$E_{hop} = E_{switch} + E_{interconnect}, \quad (4)$$

where E_{switch} and $E_{interconnect}$depend on the total capacitances and signal activity of the switch and each section of interconnect wire, respectively. They are determined as follows:

$$E_{switch} = \alpha_{switch} C_{switch} V^2, \quad (5)$$

$$E_{interconnect} = \alpha_{interconnect} C_{interconnect} V^2 \quad (6)$$

α_{switch} and $\alpha_{interconnect}$ and C_{switch} and $C_{interconnect}$ are the signal activities and the total capacitances of the switches and wire segments, respectively. V is the value of power supply. The energy dissipated in transporting a packet consisting of n flits over h hops can be calculated as:

$$\overline{E} = \frac{\sum_{i=1}^P E_{packet_i}}{P} = \frac{\sum_{i=1}^P \left(n_i \sum_{j=1}^{h_i} E_{hop,j} \right)}{P}. \quad (7)$$

The parameters switch and interconnect are those that capture the fact that the signal activities in the switches and the interconnect segments will be data-dependent, e.g., there may be long sequences of 1s or 0s that will not cause any transitions. Any of the different low-power coding techniquesaimed toreduce the number of transitions can be applied to any of the topologies described here. For the sake of simplicity and without loss of generality, we do not consider any specialized coding techniques in our analysis.

4. Simulation results and discussion

We used *ns2*simulator for the simulations about the throughput parameter [8].The applied constraints during simulation are shown in Table 1 and the correspondent results – in Fig.2 to Fig.4. Wormhole switching technique and shortest path algorithm was implemented on the different NoC topologies. The keyfactor evaluated in this case study will be the throughput for different topologies and different number of IP cores in topologies.

5. Conclusions and future work

From the simulations made with *ns 2* network simulator show one of the key performance like throughput is, of the differenttopologies of NoC's. Deep empiric investigation wasdone

NoC Model Parameters	Parameters Constraints applied in NS2
Number of Resources IP cores	16
Connections	Resource-Router, Router-Router
Transmission Proto	User Datagram Protocol(UDP)
Routing Scheme	Static
Routing Protocol	Shortest Path
Queve mechanism	Stochastic Fairness Queuing (SFQ)
Link Queue	8 packets
Bisection Bandwidth (Max.)	Router-to-router-300Mb Resource-to-router - 200Mb
Traffic Generation	Constant Bit Rate (CBR)
Traffic Rate	180 Mb
Packet Size	16 bytes

Table1. Constraints applied in *ns2* to simulate the NoC's

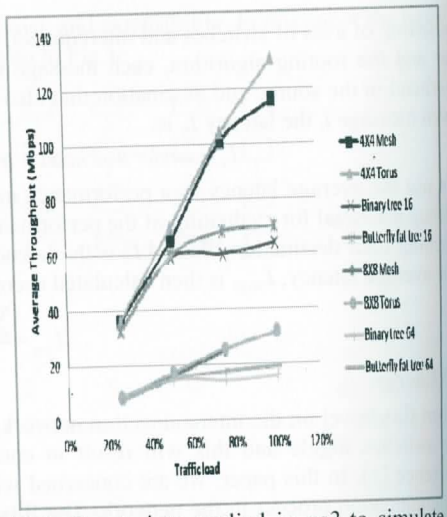


Fig.2Average throughput for different topologies and number of IP cores

Average throughput (Mbps)				
Load	4X4 Mesh	4X4 Torus	Binary tree	Butterfly Fat Tree
25%	35.945	35.862	32.753	32.659
50%	65.12	69.781	58.842	59.783
75%	100.869	103.853	59.894	68.548
100%	115.934	130.964	63.792	70.158

Fig.3 Average throughput with 16 IP cores

Average throughput (Mbps)				
Load	8X8 Mesh	8X8 Torus	Binary tree	Butterfly Fat Tree
25%	8.659	8.568	8.058	8.026
50%	16.247	17.237	14.752	14.892
75%	25.178	25.632	14.293	17.451
100%	28.589	31.641	15.491	19.058

Fig.4 Average throughput with 64 IP cores

respectively to the performance of NOC's. For future work we plan to work in the improvement of the performance empiric equations. Main direction will be reducing of the consumed energy for transfer of single flit and improvement of the existing routing algorithms to achieve minimal latency and maximal throughput. Another important direction of research is area that will be occupied on silicon slice by the Ethernet smart switch.

6. References

1. P.Pande, C.Grecu, *et al*, Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures" IEEE Trans. on Computers, v. 54, no. 8, Aug.2005;
2. C.Grecu, A.Ivanov, R.Saleh, P.Pande, Testing Network-on-Chip Communication Fabrics, IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, 2007.
3. T.Reddy, J.Singh, K.Mahapatra, Performance assessment of different NoC topologies, 2nd International Conference on Devices, Circuits and Systems (ICDCS), pp.1-5, 2014.
4. L. Benini,and G. DeMicheli, Networks on Chips: A New SoCParadigm, Computer, vol. 35, no. 1, pp. 70-78, Jan. 2002.
5. P. Magarshack and P.G. Paulin, System-on-Chip beyond theNanometer Wall, Proc. Design Automation Conf. (DAC), pp. 419-424, June 2003.
6. M. Horowitz and B. Dally, How Scaling Will Change Processor Architecture, Proc. Int. Solid State Circuits Conf. (ISSCC), pp. 132-133, Feb. 2004.
7. P. Pande, C. Grecu, *et al*, Design of a Switch for Network on Chip Applications, Proc. Int. Symp. Circuits and Systems (ISCAS), vol. 5, pp. 217-220, May 2003.
8. ns 2 website [Online]. Available: http://www.isi.edu/nsnam/ns