A Hierarchical Hybrid Optical-Electronic Network-on-Chip

Kwai Hung Mo[1], Yaoyao Ye[1], Xiaowen Wu[1], Wei Zhang[2], Weichen Liu[1], Jiang Xu[1]
1. ECE, Hong Kong University of Science and Technology
2. School of Computer Engineering, Nanyang Technological University
zhangwei@ntu.edu.sg, jiang.xu@ust.hk

Abstract— **Network-on-chip (NoC) can improve the performance, power efficiency, and scalability of multiprocessor system-on-chip (MPSoC). However, traditional NoCs using metallic interconnects consume significant amount of power to deliver even higher communication bandwidth required in the near future. Optical NoCs are based on CMOS-compatible optical waveguides and microresonators, and promise significant bandwidth and power advantages. In this paper, we propose a hybrid optical mesh NoC, HOME, which utilizes optical waveguides as well as metallic interconnects in a hierarchical manner. HOME employs a new set of protocols to improve the network throughput and latency. We compared HOME with a matched optical mesh NoC for a 64-core MPSoC in 45nm, using SPICE simulations and our cycle-accurate multi-objective NoC simulation platform, MoLab. Comparing with the optical mesh NoC, HOME uses 75% less optical/electronic interfaces and laser diodes. Simulation results show that HOME achieves 17% higher throughput and 40% less latency while consuming 42% less power.**

Keywords - hierarchical architecture; hybrid, optical; optical network-on-chip

I. INTRODUCTION

Multiprocessor system-on-chip (MPSoC) has become an important approach to achieve high performance and low power consumption by providing parallelization among a large number of functional units [1][2]. Due to technology scaling and increasing number of Intellectual Property (IP) cores in MPSoC, traditional on-chip communication architectures are no longer suitable because of the limited scalability and throughput and high power consumption. To solve this problem, network-on-chip (NoC) is introduced to facilitate the on-chip communications [1][3][4]. By placing the IP cores which communicate more frequently closer to each other, we could increase the spatial locality of the network traffic and reduce global communications [5][6].

International Technology Roadmap for Semiconductors (ITRS) shows that on-chip local clock speed could achieve 73GHz in 2020, and the limited bandwidth and high power consumption of metallic interconnect system will become a bottleneck for MPSoC [7]. Supported by recent advances in fabrication and integration of on-chip CMOS-compatible optical components [8], optical NoC promises high bandwidth, low transmission latency, and low power consumption. An optical NoC consists of optical routers, optical waveguides and IP cores. Microresonators (MRs) are used in the optical router to perform the switching function [9][10].

Several optical NoCs using optical waveguides and MRs have been proposed. Shacham *et al*. proposed a NoC system which uses an optical network for large packet transmission and an electronic network for both the control data and small packets [11]. A novel optical NoC architecture, Corona was proposed to use a pure optical arbitration scheme [12]. Batten *et al*. proposed an optical NoC with the nodes connected in a mesh with global crossbar topology [13], where optical interconnect is used for the high throughput traffic and metallic interconnect for local and fast switching. Gu *et al*. proposed a fat tree-based optical NoC [14]. Kirman *et al*. proposed an opto-electrical system [15], in which processors and memory system are connected to a switch with electronic links and the switches are interconnected using an optical ring with wavelength division multiplexing (WDM) support. A non-blocking switch structure called λ router was proposed by Briere *et al*. [16], it uses multiple 2x2 switching elements with WDM support to form a multi-stage passive photonic routing structure.

In this paper, we propose a hybrid optical mesh NoC, HOME, which utilizes optical waveguides as well as metallic interconnects in a hierarchical manner. HOME employs a new set of protocols which improve network throughput and reduce latency. HOME uses hybrid optical-electronic routers for wormhole switching in local networks and circuit switching in the global network. Packets transmit in a local network only involve a local electronic switching fabric, which minimizes the number of optical and electronic components involved and significantly reduces both power consumption and latency. For the long-distance communications among different clusters, optical circuit switching is used to achieve high bandwidth and low transmission delay with low power consumption. We compared HOME with a matched optical mesh NoC for a 64-core MPSoC in 45nm, using SPICE simulations and our cycle-accurate multi-objective NoC simulation platform, MoLab. Simulation results show that HOME uses 75% less optical/electronic interfaces and laser diodes, and consumes 42% less power, while achieving 17% higher throughput and 40% less latency.

IEEE
computer
society

The rest of the paper is organized as follows. Section II details the HOME architecture and its router. The HOME protocols used in local electronic wormhole switching clusters and global optical circuit switching network are described in Section III. Section IV shows the simulation results, comparison and analysis of throughput, latency and power consumption. Conclusions are drawn in Section V.

Figure 1. HOME for a 64-core MPSoC

II. HOME ARCHITECTURE AND ROUTER

HOME is a hierarchical optical NoC which connects MxN processor cores with hybrid optical-electronic routers (Fig. 1). The hierarchical structure helps to reduce network contentions due to interferences among short-distance and long-distance communication traffic. Four cores are connected to a HOME router using metallic interconnects to form a cluster. HOME routers are interconnected using optical interconnects to form an $\frac{M}{2} * \frac{N}{2}$ optical mesh for payload data packet. An $\frac{M}{2} * \frac{N}{2}$ electronic mesh serves as a control network to transmit control packets for maintaining optical paths. Each core is assigned a unique main ID and sub ID pair for addressing. The main ID is used to identify a cluster, and the sub ID is used to identify a core within a cluster. The optical routers and interconnects can be fabricated on different device layers using 3D integration technology to reduce the chip size and facilitate different fabrication processes [17].

Figure 2. HOME router

HOME router is a hybrid optical-electronic router, consisting of an optical switching fabric, an electronic switching fabric, a router control unit (RCU) and an optical/electronic (O/E) interface (Fig. 2). The hybrid router architecture reduces the interconnect length and power consumption for intra-cluster communications. Both the optical

switching fabric and electronic switching fabric are controlled by the RCU. Communications between electronic and optical domain are facilitated by the O/E interface, which handles serialization, deserialization and O/E conversions. The 5x5 electronic switching fabric is composed of five input buffers and a 5x5 crossbar (Fig. 3). Four ports are connected to local processor cores, and one is connected to the O/E interface. The 5x5 non-blocking crossbar allows five concurrent transactions if there is no contention for the same output port.

RCU implements the routing algorithm and protocols and controls the switching fabrics. For intra-cluster communications, only the electronic switching fabric is used and no global traffic will be generated. For inter-cluster communications, payload data packets will pass the electronic switching fabric and use the optical switch fabric. RCU uses optical circuit switching for inter-cluster data packets. It reserves an optical path between the source and destination cores prior to payload data packet transmissions. The setup, maintenance and teardown of an optical path are achieved by sending control packets through the electronic control network. RCU uses electrical signals to control the MRs inside the optical switching fabric.

Figure 3. Electronic switching fabric

The optical switching fabric is based on two types of basic 1x2 optical switching elements -- crossing element and parallel element, which both consist of one MR and two waveguides (Fig. 4). The crossing element has an optical waveguide crossing inside and hence extra optical power loss compared with the parallel element. In the basic switching elements, when powered on, the MRs have an on-state resonance wavelength λ_{on}. While an optical signal with a centre wavelength λ_{on} comes into the input port, a powered-on MR will deliver the optical signal to the drop port; otherwise, the optical signal will be delivered to the through port. As a result, the basic elements achieve 1x2 optical switching functions by powering on or off the MRs.

Figure 4. Basic 1x2 optical switching elements

The 5x5 optical switching fabric is strictly non-blocking (Fig. 5). It has 16 MRs, six waveguides and two optical

terminators. The optical switching fabric is designed to minimize the number of waveguide crossings, MRs and MR switching activities. There are five bidirectional optical ports, including East, South, West, North and a local port (injection/ejection) which is connected to the O/E interface. The ports are aligned to their intended directions so no extra crossings are incurred at network level. The optical switching fabric has a special feature. It does not power on any MR for the same dimension transmissions, i.e. from south to north, east to west and vice-versa.
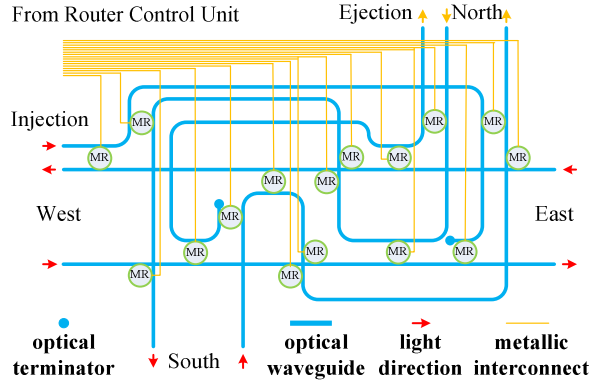
Figure 5. Optical switching fabric

III. HOME PROTOCOLS

HOME uses wormhole switching for intra-cluster communications. Routing information, including the source address, destination address and packet size, is stored in the first flit of a packet called header flit, which is followed by payload flits. RCU exams the routing information in the header flit and makes routing decisions. If the main ID of the destination core matches the main ID of the current cluster, which means the source and destination cores are in the same cluster, the header flit and subsequent payload flits will be forwarded to the corresponding port. If the port is blocked, the packet will be held until the port becomes available. Round-robin arbitration is used to solve port contentions. No interrupt is allowed before a packet transfer is completed. If the source and destination cores are in different clusters, a global optical transmission will be initiated. Due to the difficulties in buffering optical signals, circuit switching is used to establish a connection-oriented optical path between the source and destination cores. HOME uses the electronic control network to setup, maintain and tear down optical paths. The electronic control network is formed by connecting RCUs in an $\frac{M}{2} * \frac{N}{2}$ mesh topology. It uses dimension order routing, which is an oblivious and deterministic routing algorithm. Dimension order routing is deadlock-free and supports in-order packet delivery.

A new set of protocols are developed to reduce latency and increase network throughput in optical NoCs. There are four phases for optical data transmissions, including optical link reservation, acknowledgement, data transmission and path teardown. When RCU finds a packet with the destination main ID belongs to another cluster, it will generate a path setup packet to reserve the optical links for payload packet transmission. Every time the path setup packet progressed to

the next router, the optical waveguide and electronic interconnect of the control network between the previous and current router will be reserved and the optical switching fabric of the routers are configured. After the path setup packet reaches the RCU of the destination cluster, an Ack packet will be sent from the destination RCU to the source RCU when the destination port connected to the targeted core is available. In order to reduce the path setup latency, the Ack packet is transmitted along the reserved optical path based on the optical symmetry of HOME. As a result, the Ack packet can reach the source RCU quickly. This also eliminates the need to hold a bidirectional control path associated with the optical path in the control network as well as the reverse-direction optical path associated with the reserved optical path. When the source RCU receives the Ack packet, optical data transmission can begin. The packet from the source core will pass the electronic switching fabric to O/E interface, which then converts the packet into optical signal and injects it to the injection port of the optical switching fabric. At the destination cluster, the packet is delivered by a reversed procedure.

After data transmission, the reserved optical path should be released. Traditionally, after the optical data transmission, a teardown control packet will be sent from the source to the destination through the electronic control network in order to inform the intermediate routers to release any reserved resource associated with the optical transmission. It takes a large number of clock cycles to route the teardown packet from the source to destination, which takes a substantial amount of time and reduces the availability of network resource. We propose a new method to reduce teardown time. The teardown packet will be sent to the destination RCU at the same time when the optical data transmission begins. After the optical path is reserved, with known transmitted packet size L_{packet}, optical interconnect bandwidth $B_{optical}$ and control network frequency $f_{control}$, we can easily calculate the number of cycles $C_{transmission}$ required for the optical transmission by (1).

$$C_{transmission} = \left\lceil \frac{L_{packet} \cdot f_{control}}{B_{optical}} \right\rceil \qquad (1)$$

$C_{transmission}$ is stored in the time-to-live (TTL) field of the teardown control packet. There is a countdown counter in each RCU which will be set to the TTL value and starts the countdown immediately when RCU receives a teardown packet. When the counter reaches zero, it means the optical packet transmission is completed and then the RCU will release the reserved resources associated with that optical transmission. Each time when the RCU forwards this teardown packet to the next hop, the TTL field will be reduced by the elapsed number of clock cycles for travelling one hop. If the optical data transmission time is less than the time required for the control packet to reach the destination, the TTL value of the teardown packet will reach zero in the middle. An intermediate RCU receiving a teardown packet with a zero value in the TTL field will release the resource immediately.

IV. COMPARISON AND ANALYSIS

We analyzed the power consumption, network throughput, and latency of HOME for a 64-core MPSoC in 45nm, and compare it with a matched optical mesh NoC under different packet sizes, offered loads and traffic patterns. The analysis and comparison are based on SPICE simulations as well as our cycle-accurate multi-objective NoC simulation platform, MoLab. MoLab is based on NS-2 [18]. It can model both optical and electronic NoCs, and analyze network throughput, latency, power consumption, average hop distance, etc, under different communication scenarios and traffic patterns.

In the electronic switching fabrics of HOME routers, each port is 32-bit width and operates at 1.25GHz. The input buffer in each port is 2-flit deep. In the optical switching fabrics of HOME routers, each port has a bandwidth of 40Gbps using wavelength division multiplex (WDM) technique. A matched optical mesh NoC is designed using exactly the same parameters. Compared to the optical mesh, HOME uses 75% less O/E interfaces and laser diodes, which significantly reduces the cost of HOME.

A. Network Throughput and Latency

The performance of a network is measured in terms of throughput and average packet latency. Packet latency is the time between the first flit of a packet generated and the last flit of the packet received by its destination. Packets are generated by each core and have various destinations. We model the packet destinations as a Gaussian distribution in (2), where μ is the mean, σ is the standard deviation (SD).

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{2}$$

The injection rate α is calculated by (3), where T_{busy} is the time to transmit packets and T_{idle} is the average time interval between generating two successive packets. T_{idle} follows an exponential distribution showed in (4), where λ is the mean and equals to T_{idle}.

$$\alpha = \frac{T_{busy}}{T_{busy} + T_{idle}} \tag{3}$$

$$p(x;\lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{1}{\lambda}x}, x \geq 0 \\ 0, \qquad x < 0 \end{cases} \tag{4}$$

Network throughputs and average packet latencies under different packet sizes and traffic patterns are shown in Fig. 6 to 8. Before the saturation point, the network throughput is almost linear to the injection rate. When the injection rate reaches the saturation point, the network is overloaded, and the packets begin to accumulate at the source, which causes the latency to rise significantly and the network throughput to reach its maximum level. The larger the SD is, the more packets have a destination far from the source. For example, when SD is equal to two, about 46% packets are inter-cluster traffic. As packets need to traverse more routers before reaching destinations, the average packet latency increases while the saturation point

decreases. The network throughput is also reduced due to the increasing transmission time. The network throughput increases along with the growing packet size. Analysis shows that, for a fixed injection rate, a larger packet size can reduce packet number and increase packet interval, which in turn reduces the routing overheads and network contentions. The saturation point is also increased along with the growing packet size due to the path setup latency is amortized by larger packet sizes.

Simulations and analysis show that HOME has better network throughput and average packet latency than the optical matched mesh NoC by using the hierarchical hybrid architecture and the new protocols. The average packet latency of the optical mesh NoC is 15.4ns before the saturation point for 128-bit packets while SD is equal to two. In comparison, HOME reduces 40% average packet latency which is only 9.2ns before the saturation point. The network throughput of the optical mesh NoC is 340Gbps for 128-bit packets while SD is equal to two. In comparison, HOME increases 17% network throughput to 399Gbps.

B. Power Consumption

Power consumption is a critical aspect of NoC design. For high-performance computing, low power consumption can reduce the cost related with packaging, cooling solution, and system integration. We use MoLab to analyze the power consumption of HOME and the matched optical mesh NoC based on an analytical model.

For the inter-cluster communications in HOME, the power consumption of each packet E_{pkt}^g is calculated by (5).

$$E_{pkt}^g = E_{payload} + E_{ctrl} \tag{5}$$

$E_{payload}$ is the energy consumed to transmit the payload data and $E_{control}$ is the energy consumed by the associated control packets. $E_{payload}$ is calculated by (6).

$$E_{payload} = m \cdot P_{mr} \cdot (\frac{L_{payload}}{B} + \frac{d \cdot n}{c} + T_{setup} + T_{teardown}) \\ + E_{oeeo} \cdot L_{payload} + 2 \cdot (E_{xbar} + E_{bf} + E_{local\_int}) \cdot L_{pkt} \tag{6}$$

m is the number of MRs in the on-state while transferring the payload packet, P_{mr} is the average power consumed by a MR when it is in the on-state. $L_{payload}$ is the payload packet size. B is the communication bandwidth. d is the distance traveled by the payload packet. c is the speed of light in vacuum. n is the refractive index of optical waveguide. E_{oeeo} is the energy consumed for 1-bit OE and EO conversions. T_{setup} is the optical path setup time. $T_{teardown}$ is the optical path teardown time. E_{xbar} is the average energy required to transfer a single bit through a crossbar. E_{local_int} is the average energy required to transfer a single bit through an metallic interconnect between a core and a router. E_{bf} is the average energy required for the buffer to store a single bit. L_{pkt} is the packet size in the electronic NoC. $E_{control}$ is calculated by (7), where E_{global_int} is the average energy required to transfer a single bit through a metallic interconnect between routers, E_{rcu} is the average energy required by the

control unit to make decisions for a single packet, $L_{control}$ is the control packet size, and h is the number of hops.

$$E_{ctrl} = E_{global\_int} \cdot h \cdot L_{ctl} + E_{rcu} \cdot (h+1) \qquad (7)$$
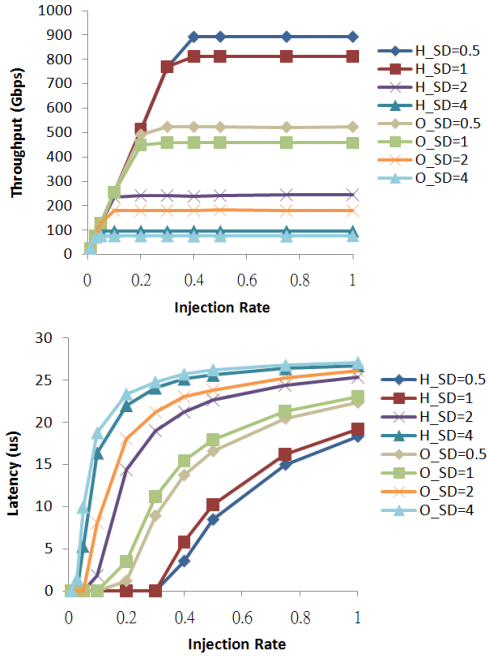

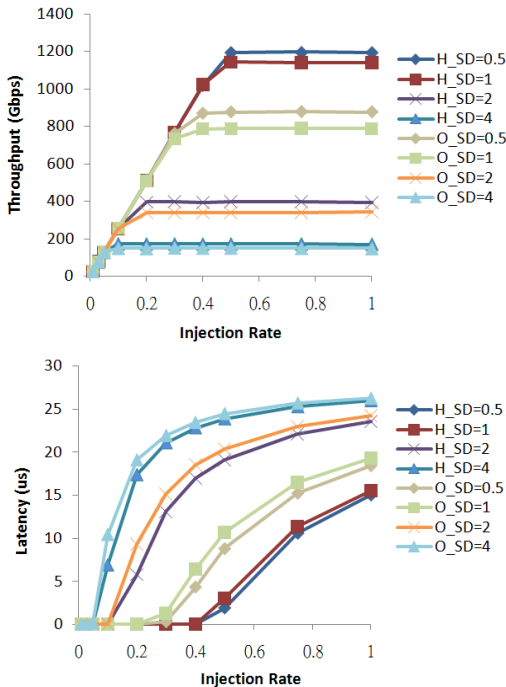Figure 6. Network performance of HOME and 8x8 optical mesh NoC for the 64-core MPSoC using 64-bit packets

Figure 7. Network performance of HOME and 8x8 optical mesh NoC for the 64-core MPSoC using 128-bit packets

For the intra-cluster communications in HOME, the power consumption of each packet E_{pkt}^{l} is calculated by (8).

$$E_{pkt}^{l} = (E_{xbar} + E_{bf} + 2 \cdot E_{local\_int}) \cdot L_{pkt} + E_{rcu} \qquad (8)$$

We designed and simulated the 5x5 input-buffered electronic switching fabric and RCU based on the 45nm Nangate open cell library and Predictive Technology Model [19] in Cadence Spectre. We assume each processor core is 1mm by 1mm. The metal wires in the electronic NoC are modeled as fine-grained lumped RLC networks. Since mutual inductance has a significant effect in deep submicron process technologies, the coupling capacitances among adjacent wires is considered up to the third neighboring wires. Simulation results show that on average the crossbar of the router consumes 0.07pJ/bit, the input buffer consumes 0.003pJ/bit, and the RCU consumes 1.8pJ to make decisions for each packet. For the metal wires, global interconnect takes 0.62pJ/bit for the communication between routers and local interconnect between router and processor takes 0.04pJ/bit. The energy consumed by OE and EO conversions is estimated to be 1 pJ/bit, which is scaled down from an 80nm design [20] to 45nm. Each MR needs a DC current and consumes less than 20μW in the on state [21].
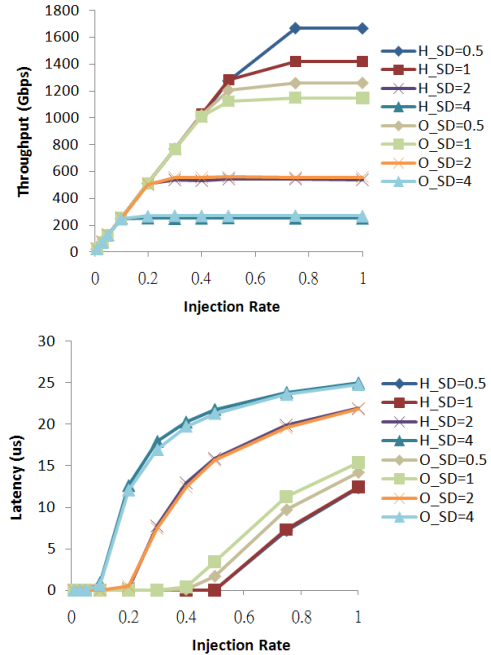

Figure 8. Network performance of HOME and 8x8 optical mesh NoC for the 64-core MPSoC using 256-bit packets

Simulations and analysis results show that HOME has better power efficiency compared with the matched optical mesh NoC (Fig. 9). For example, while the optical mesh NoC consumes 1.5pJ/bit to transmit 128-bit packets when SD is two, HOME only consumes 0.86pJ/bit, which is 42% lower. HOME achieves lower power consumption by using less number of hops for long-distance communications and only the electronic switching fabrics for local communications. Since local communications only involve the local router, this eliminates

the energy dissipation on global interconnects and minimizes the components involved. When SD increases, more packets will traverse larger distance, and the power consumption increases. However, when the packet size increases, the control overhead will be reduced, and this helps to lower the power consumption.
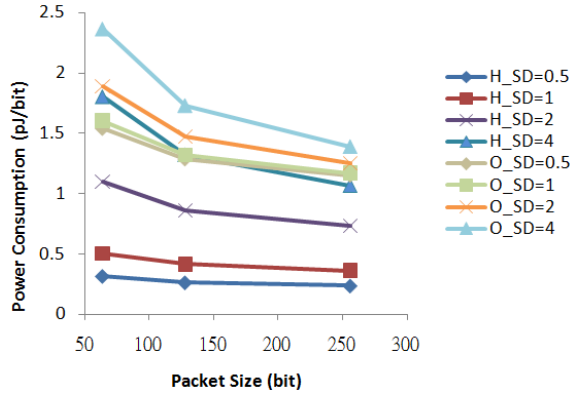

Figure 9. Power consumption of HOME and the 8x8 optical mesh NoC for the 64-core MPSoC

V. CONCLUSIONS

This paper proposes a hybrid optical mesh NoC, HOME. HOME utilizes optical waveguides as well as metallic interconnects in a hierarchical manner. It employs a new set of protocols which can improve network throughput and reduce latency. HOME uses hybrid optical-electronic routers for wormhole switching in local networks and circuit switching in the global network. For the long-distance communications among different clusters, optical circuit switching is used to achieve high bandwidth, low transmission delay, and low power consumption. We compared HOME with a matched optical mesh NoC for a 64-core MPSoC in 45nm, using SPICE simulations and our cycle-accurate multi-objective NoC simulation platform, MoLab. HOME uses 75% less optical/electronic interfaces and laser diodes. Simulation results show that HOME achieves 17% higher throughput and 40% less latency while consuming 42% less power.

REFERENCES

[1] L.Benini and G. De Micheli, "Networks on chip: a new paradigm for system on chip design," *Design, Automation and Test in Europe conference*, 2002.

[2] K. Lahiri, A. Raghunathan, G. Lakshminarayana and S. Dey, "Design of high-performance System-on-chips using Communication Architecture Tuners," *IEEE Transactions on Computer-Aided Design*, 2004.

[3] J. Xu, W. Wolf, J. Henkel, S. Chakradhar, "A Design Methodology for Application-Specific Networks-on-Chip", *ACM Transactions in Embedded Computing Systems*, July 2006.

[4] A. Hansson, K. Goossens and A. Radulescu, "A Unified Approach to Mapping and Routing on a Network on Chip for both Best-Effort and Guaranteed Services," *VLSI Design, Special Issue on Networks on Chip*, 2007.

[5] J. Hu and R. Marculescu, "Energy- and Performance-Aware Mapping for Regular NoC Architectures," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2005

[6] W.Zhou, Y. Zhang and Z.Mao, "An Application Specific NoC Mapping for Optimized Delay," *IEEE conference on Design and Test of Integrated Systems in Nanoscale Technology*, 2006.

[7] R. G. Beausoleil, P. J. Kuekes, G. S. Snider, S.-H. Wang, and R. S. Williams, "Nanoelectronic and nanophotonic interconnect," *Proceedings of the IEEE*, 2008.

[8] G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet and E. G. Friedman, "Predictions of CMOS Compatible On-Chip Optical Interconnect," *ACM/IEEE International Workshop on System Level Interconnect Prediction*, 2005.

[9] B. E. Little, H. A. Haus, J. S. Foresi, L. C. Kimerling, E. P. Ippen, and D. J. Ripin, "Wavelength switching and routing using absorption and resonance," *Photonics Technology Letters*, 1998.

[10] B. E. Little, J. S. Foresi, G. Steinmeyer, E. R. Thoen, S. T. Chu, H. A. Haus, E. P. Ippen, L. C. Kimerling, and W. Greene, "Ultra-compact Si-SiO2 microring resonator optical channel dropping filters," *IEEE Photonics Technology Letters*, 1998

[11] A. Shacham, B.G. Lee, A. Biberman, K. Bergman, and L.P. Carloni, "Photonic NoC for DMA Communications in Chip Multiprocessors," *Hot Interconnects*, 2007.

[12] D. Vantrease, R. Schreiber, M. Monchiero, *et al*, "Corona: System Implications of Emerging Nanophotonic Technology," *International Symposium on Computer Architecture*, 2008.

[13] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. Holzwarth, M. Popovic, L. Hanqing, H. Smith, J. Hoyt, F. Kartner, R. Ram, V. Stojanovic, and K. Asanovic, "Building Manycore Processor-to-DRAM Networks with Monolithic Silicon Photonics," *IEEE Symposium on High Performance Interconnects*, 2008.

[14] H. Gu, J. Xu, W. Zhang, "A Low-power Fat Tree-based Optical Network-on-Chip for Multiprocessor System-on-Chip", *Design, Automation and Test in Europe Conference and Exhibition*, 2009.

[15] N. Kirman, M. Kirman, R. K. Dokania, J. F. Martinez, A. B. Apsel, M. A. Watkins, and D. H. Albonesi, "Leveraging Optical Technology in Future Bus-based Chip Multiprocessors," *Annual IEEE/ACM International Symposium on Microarchitecture*, 2006.

[16] M. Briere, B. Girodias, et al, "System Level Assessment of an Optical NoC in an MPSoC Platform," *Design, Automation & Test in Europe Conference & Exhibition*, 2007.

[17] J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, N. Vijaykrishnan, C. Das, "A Novel Dimensionally-Decomposed Router for On-Chip Communication in 3D Architectures," *the Annual International Symposium on Computer Architecture*, 2007.

[18] http://nsnam.isi.edu/nsnam

[19] www.si2.org

[20] C. Kromer, G. Sialm, C. Berger, T. Morf, M.L. Schmatz, F. Ellinger, et al., "A 100-mW 4×10 Gb/s transceiver in 80-nm CMOS for high-density optical interconnects," *IEEE Journal of Solid-State Circuit*, 2005.

[21] A. W. Poon, F. Xu, X. Luo, "Cascaded active silicon microresonator array cross-connect circuits for WDM networks-on-chip," *Proceedings of SPIE*, 2008.