Exploring Hybrid Photonic Networks-on-Chip for Emerging Chip Multiprocessors

Shirish Bahirat, Sudeep Pasricha
Department of Electrical and Computer Engineering
Colorado State University, Fort Collins, CO
{shirish.bahirat, sudeep.pasricha}@colostate.edu

ABSTRACT

Increasing application complexity and improvements in process technology have today enabled chip multiprocessors (CMPs) with tens to hundreds of cores on a chip. Networks on Chip (NoCs) have emerged as scalable communication fabrics that can support high bandwidths for these massively parallel systems. However, traditional electrical NoC implementations still need to overcome the challenges of high data transfer latencies and large power consumption. On-chip photonic interconnects have recently been proposed as an alternative to address these challenges, with high performance-per-watt characteristics for intra-chip communication. In this paper, we explore using photonic interconnects on a chip to enhance traditional electrical NoCs. Our proposed hybrid photonic NoC utilizes a photonic ring waveguide to enhance a traditional 2D electrical mesh NoC. Experimental results indicate a strong motivation for considering the proposed hybrid photonic NoC for future CMPs – as much as a 13× reduction in power consumption and improved throughput and access latencies, compared to traditional electrical 2D mesh and torus NoC architectures.

Categories and Subject Descriptors: C.5.4 [**VLSI Systems**]
General Terms: Design, Experimentation, Power, Performance.
Keywords: Network-on-chip, photonic interconnect, Chip Multiprocessor.

1. INTRODUCTION

Driven by increasing application complexity and improvements in fabrication technology, chip multiprocessors (CMPs) with tens to hundreds of processing cores on a chip are becoming increasingly prevalent [1]-[3]. In order for cores to communicate efficiently with each other, an efficient on-chip communication fabric is essential. Consequently, over the last few years, networks-on-chip (NoCs) have received much research interest, for CMPs as well as application specific multi-processor systems-on-chip (MPSoCs) [4]. NoCs offer significant benefits in bandwidth, scalability, and reliability compared to traditional hierarchical and crossbar-based shared bus communication architectures in ultra deep submicron (UDSM) technologies [5].

Practically, NoC communication fabrics must still overcome two major challenges that are extremely relevant for emerging CMP applications [6]. Firstly, NoCs must enable low latency transfers between cores. Packet switched networks are often unable to meet quality of service guarantees, which makes it difficult to provide bounds on packet latency, essential especially for applications with real time constraints. Using virtual circuit switching provides a better alternative in such a scenario and generally enables lower latency transfers than packet switching. But there is still an inherently unpredictable delay in setting up the circuit, lower resource utilization, and much lower performance for non-circuit switched streams. Secondly, NoCs must enable low power data transfers. However, the large number of network interface, routers, links, and buffers that are part of the NoC fabric lead to the communication infrastructure consuming significant power (as much as 25% of the overall CMP power budget [7]). This power consumption becomes worse when the NoC components are designed for adaptive operation to cope with varying application requirements at runtime. Since power consumption has a major impact on maximum temperature which determines packaging and cooling costs, and NoC power consumption will only increase as the number of on-chip cores increase, reducing NoC power consumption becomes vitally important. Recent studies highlight this phenomenon, indicating that NoC power consumption based on current circuit implementation techniques is much higher (a factor of 10×) than what is needed to meet the expected needs of future CMPs [8].

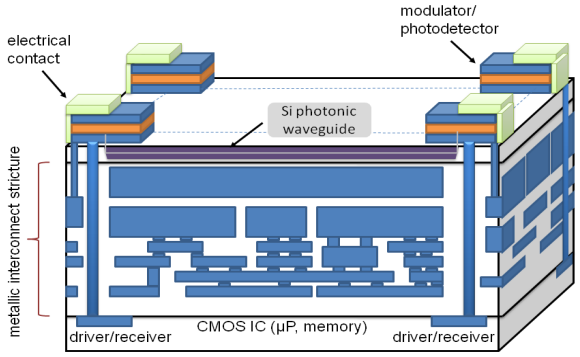


Figure 1. 3D implementation of the proposed hybrid photonic NoC - mesh of cores on the bottom layer, photonic waveguide on the top layer

A promising recent research direction in order to overcome the abovementioned drawbacks is to make use of photonic communication on a chip [9][10]. Such photonic interconnects can theoretically offer ultra-high communications bandwidths in the terabits per second range, in addition to lower access latency and susceptibility to electromagnetic interference [11]. Photonic signaling also has low power consumption, since the power consumption of optically transmitted signals at the chip level is independent of the distance covered by the optical signal [12]. While photonic NoCs were virtually inconceivable with previous generations of photonic technologies a few years ago, recent

CODES+ISSS'09, October 11–16, 2009, Grenoble, France.

advances in the field of nanoscale silicon (Si) photonics have enabled the possibility of creating highly integrated photonic CMOS NoC platforms that can send and receive optical signals with superior power efficiencies [13]-[17]. In fact, photonic elements have recently become available as library cells in standard CMOS processes. It has thus now become practical to realistically consider an interconnection network for CMPs built with photonic elements.

In order to implement a photonic NoC, it is highly likely that future CMOS ICs will utilize 3D integration [18] as shown conceptually in Fig. 1. 3D integration will allow logic and silicon (Si) photonics planes to be separately optimized [9][53]. In the figure, the bottom plane consists of a CMOS IC with several microprocessor and memory cores, while the top plane consists of the photonic waveguide. It is also possible for all memory cores to be implemented on a dedicated layer, separate from the microprocessor layer. Vertical interconnects to connect these planes are accomplished with electrical through silicon vias (TSVs). Since photonic memories and photonic packet routing cannot be easily implemented in Si, an electrical portion of the NoC is responsible for interfacing with and controlling transfers to and from processor and memory cores via the photonic path.

Based on the recent advances in integrating nanoscale silicon photonics with commercial CMOS manufacturing technology, in this paper we explore a novel hybrid electro-photonic NoC fabric for emerging CMP applications. Our proposed fabric consists of a photonic ring waveguide that acts as a global communication channel and complements a more traditional 2D electrical NoC fabric. This hybrid communication architecture utilizes electrical and photonic paths simultaneously to improve performance per watt characteristics for a CMP. We explore different architectural configurations of our hybrid photonic NoC fabric by considering *(i)* varying levels of electrical to photonic communication connectivity, and *(ii)* multiple degrees of communication serialization. These configurations enable interesting tradeoffs between performance and power consumption of the proposed architecture. Our experimental results indicate significant potential for using the proposed hybrid photonic NoC communication fabric for emerging CMPs – as much as a 13× reduction in power consumption and improved throughput and access latencies, compared to traditional electrical 2D mesh and torus NoCs.

2. RELATED WORK

The concept of photonic interconnects for on-chip communication was first introduced by Goodman et al. [19]. Since then, with advances in the fabrication and integration of photonic elements on a CMOS chip, several works have presented a comparison of the physical properties of on-chip electrical (copper-based) and photonic interconnects [20]-[26]. In particular, Collet et al. [20] compared simple photonic and point-to-point links with a Spice-like simulator. Tosik et al. [22] studied more complex interconnects, comparing photonic and electrical clock distribution networks, using accurate physical simulations, synthesis techniques and predictive transistor models. Both works studied power consumption and bandwidth, and highlighted the benefits of on-chip optical interconnect technology. Intel's Technology and Manufacturing Group also performed a preliminary study evaluating the benefits of optical intra-chip interconnects [23]. They concluded that while optical clock distribution networks are not especially attractive, wave division multiplexing (WDM) based on-chip optical interconnects offer interesting advantages for intra-chip over copper in UDSM process technologies.

In addition to comparisons between photonic and electrical interconnects at the physical level, a few recent works have explored the system level impact of using photonic interconnects and explored photonic NoCs [4][10][27]-[31]. Our previous work [27] explored a photonic ring interconnect to improve the latency response and reduce power consumption of circuit switched bus-based communication architectures for application specific MPSoCs. Briere et al. [28] proposed a crossbar based photonic NoC, comprised of passive resonator devices with routing between an input output pair being achieved by selecting the appropriate wavelength. Vantrease et al. [29] proposed a similar dense WDM based optical crossbar for on-chip communication. Shacham et al. [10][30] explored the feasibility of a hybrid optical NoC where a photonic NoC comprised of silicon photonic switches connected by waveguides is used to transmit large messages, and an electronic control network, topologically identical to the photonic network, is "folded" within the photonic network to control its operations and execute exchange of short messages. A few other works have also explored non-blocking micro-resonator based photonic switches [4][28][31] for routing photonic messages.

Unlike the works discussed above, in this paper we explore a novel photonic ring based hybrid photonic NoC that augments a traditional 2D all-electrical mesh NoC. The photonic ring is used primarily to facilitate global on-chip communication between distant processor and memory cores on the chip. The simple ring waveguide dramatically reduces photonic path complexity, compared to some of the works discussed above that propose complex photonic crossbars, meshes and tori, while still providing significant opportunity for improvements over traditional, all-electrical NoCs. We explore several different configurations of our proposed hybrid photonic NoC to analyze trade-offs between performance (throughput, latency) and power consumption for small to large CMPs.

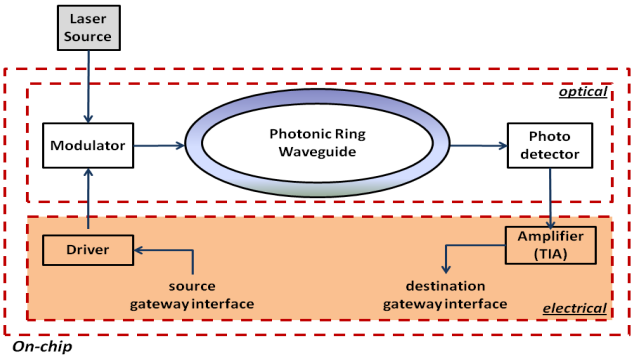


Figure 2. High level overview of photonic transmission architecture

3. HYBRID PHOTONIC NOC: OVERVIEW

Even though electronic NoCs have many advantages in terms of flexibility and functionality, they tend to consume high power, and have high access latencies, both of which scale up with the transmitted bandwidth and CMP/mesh dimension size. On the other hand, photonic technology offers significant and unique advantages in terms of power consumption, access latencies and high bandwidth. However two essential enablers for packet switching, *(i)* buffering and *(ii)* header processing are relatively difficult to implement with a photonic NoC. Thus a hybrid communication infrastructure with both photonic and electrical signaling becomes necessary.

Fig. 2 shows a high level overview of the photonic transmission components in our hybrid photonic architecture. There are four

primary photonic components: a laser (light source), an opto-electric modulator/transmitter, a photonic ring waveguide and a photonic receiver. The modulator converts electrical signals into light (E/O), which is propagated through the photonic waveguide, and then detected and converted back into an electrical signal at the receiver (O/E). Unlike the rest of components, there are still significant challenges in efficiently integrating a silicon based laser on a chip. Using an off-chip laser can actually be beneficial because it leads to lower on-chip area and power consumption. Consequently, in our architecture we assume an off-chip laser from which light is coupled onto the chip using optical fibers, much like what is done in chip-to-chip optical interconnects today [32]. The next subsection provides an overview of the photonic building blocks used in this architecture.

3.1 Photonic Architecture Building Blocks

3.1.1 Photonic Waveguide

In general, photonic waveguides with highly angled structures (such as those commonly found in electrical topologies) may result in significant signal degradation. This degradation is compounded when laying out multiple waveguides for multi-bit parallel transfers on communication links. Consequently, we propose a much simpler ring like structure that is better suited to the physical characteristics of photonic waveguides. This ring with multiple parallel waveguides resides on a dedicated Si layer and encircles a large portion of the chip, with computing nodes interfacing with it using E/O transmitters and O/E receivers.

For on-chip photonic interconnects, there are two popular candidates for waveguide material: high refractive index silicon on insulator (SOI) and low refractive index polymer waveguides. SOI waveguides have lower pitch (i.e., width) and lower area footprint compared to polymer waveguides. This leads to better bandwidth density (i.e., transmitted bits per unit area). In addition, although modulators exist for both silicon [33] and polymer waveguides [34], polymer-based modulators are bulky, and require high voltage drive for high frequency operation. These drawbacks limit their applicability to on-chip photonic links. Consequently, we make use of SOI as the photonic ring waveguide material.

To support high peak bandwidths in excess of 600 Gb/s for future CMP applications, the photonic ring supports wave division multiplexing (WDM), with *8* wavelengths available for every waveguide [24]. In addition, a 12× optical time division multiplexing (OTDM) is used to multiplex the modulated data streams for each wavelength to achieve even higher transmission capacity [35]. Our WDM implementation employs multiplexing by core, with each of the *n* interfacing cores having exclusive access to *8/n* wavelengths. This limits the number of transmitters, but provides substantial power savings. The scheme allows designers to optimize light power through individual coupling-ratio tuning at detectors at design time. If reduction in photonic waveguide complexity is important, WDM can be used to reduce the number of waveguides by a factor of 8 as well (although this is left as future work). It is also necessary in the ring structure to prevent light from circulating around the ring for more than one complete cycle, otherwise older messages may interfere with new ones. To address this issue, attenuators are placed immediately before each modulator, to act as a sink for the corresponding wavelength once the signal goes full circle.

3.1.2 E/O and O/E Converters

In order to transmit data from the cores through the photonic ring, electronic to optical (E/O) conversion at the source and an optical to electrical (O/E) conversion at the destination is required. Such components that connect the electrical and photonic realms on a chip have recently become commercially available [32]. In our architecture, we propose using small footprint microring-resonator based silicon optical modulators that can support data rates up to 12.5 Gb/s [36] for E/O conversion. The modulator is driven by a series of tapered inverters (i.e., driver). The first stage consists of a minimum sized inverter. At the receiver, SiGe photo detectors [37] are used to perform O/E conversion. Wave selector filters are used for each wavelength supported by a waveguide, under WDM. Trans-impedance amplifier (TIA) circuits are used to amplify the resulting analog electrical signal to a digital voltage level. To achieve high-gain and high-speed detection, a higher analog supply voltage than the digital supply voltage may be required, which may consume higher power. We assume a TIA supply voltage that is 20 % higher than the nominal supply for our power calculations.

3.2 System Level Architecture

At the system level, the proposed hybrid photonic NoC consists of the waveguide and photonic components on a dedicated top layer, interfacing with a traditional 2D electrical mesh NoC using a gateway interface block. This gateway interface consists of both driver buffers, TIA amplifiers, circuitry for clock synchronization and recovery, as well as serialization and de-serialization if necessary (discussed in subsection 3.4).

In the 2D electrical mesh, routers have an input queued crossbar with a 4-flit buffer on each input port, and a flit width of 256 (which is also the number of parallel photonic waveguides). Each router has five I/O ports (N, S, E, W, processing core) and implements wormhole switching, along with switch-to-switch ACK/NACK flow control. XY dimension order routing is employed to simplify router and network interface (NI) design for routing on the electrical network. Gateway interfaces enhance the functionality of a small subset of these routers, which then have additional ports to the photonic plane to pass messages to/from the photonic ring waveguide. The routing functionality in these gateway interface routers is also modified so that the photonic path is chosen instead of a traditional XY route for long distance communication. If multiple requests contend for access to the photonic waveguide at a gateway interface, then the request with the furthest distance to the destination is given preference (other schemes can be used here as well). A simple ACK/NACK flow control is used on the photonic waveguide path as well, to ensure that the destination is able to accept the data to be sent (as optical buffering is not an option). Because photonic messages are transmitted without buffering, this approach can be considered to be *photonic circuit switching* that works in tandem with *electrical packet switching*. To reduce the overhead on router complexity, only four corner routers are chosen as gateway interfaces to the photonic plane. The extent of communication over the photonic waveguide is controlled by a parameterizable region of optical influence, described next.

3.3 Region of Optical Influence

Since we restrict the number of gateway interfaces to four in our architecture, the amount of photonic path utilization may reduce as the size of the CMP and number of cores on a chip rise. To ensure appropriate scaling and utilization, we propose a parameterizable region of optical influence, which refers to the number of cores around the gateway interface that can utilize the photonic path for communication. For smaller sized systems (e.g., 3×3 CMPs), limiting the number of cores interfacing with each gateway

interface to one (i.e., region of size 1) may be sufficient to offload a majority of the global communication from the electrical network. However for more complex systems (e.g., 8×8 CMPs) a larger region size may be more appropriate. Fig. 3 shows an 8×8 CMP with varying regions of optical influence at the four interface gateways. If a router falls under the region of optical influence, it is modified to additionally consider the photonic path for global communication for incoming packets. Note that while the sizes for the regions are shown as different at each gateway interface in the figure, this is for illustration purposes only, and in practice we assume a fixed sized region of optical influence for all four gateway interfaces. Our experiments in the next section explore the impact of varying this region of optical influence on NoC performance and power consumption.

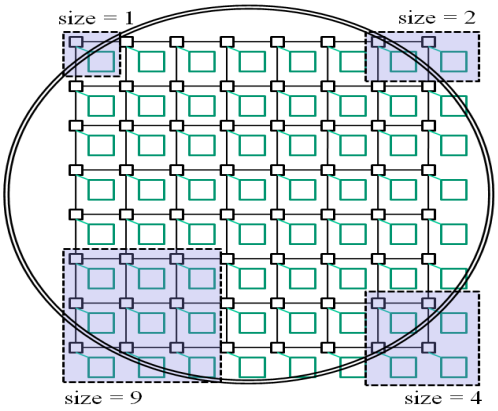


Figure 3. Regions of optical influence

3.4 Communication Serialization

Serialization of electrical communication links has been widely used in the past to reduce wiring congestion, lower power consumption (by reducing link switching and buffer resources), and improve performance (by reducing crosstalk) [38]-[40]. Since reducing power consumption is a critical design goal in future CMPs, we propose using serialization at the gateway interfaces, to reduce the number of photonic components (waveguides, buffers, transmitters/receivers), and consequently reduce area and complexity on the photonic layer as well as lower the power consumption. In our architecture, we make use of a shift register based serialization scheme, similar to [41]-[43]. A single serial line is used to communicate both data and control signals between the source and destination nodes. A frame of data transmitted on the serial line using this scheme consists of n+2 bits, which includes a start bit ('1'), n bits of data, and a stop bit ('0'). More information on the serialization scheme implementation can be found in our detailed technical report [52].

4. EXPERIMENTAL RESULTS

In this section, we present experimental results to evaluate our proposed hybrid photonic NoC architecture. The first subsection presents the estimation models used. The subsequent subsections present our various experimental results.

4.1 Estimation Models

To obtain the delay for the waveguide and other photonic components, we used results from [24][44], shown in Table 1 for the 65 nm process node. The delay of an optimally repeated and sized electrical (Cu) wire at 65 nm was assumed to be 26 ps/mm [21]. The operating frequency of the photonic ring is estimated by calculating the time needed for the light to travel from any node to the farthest node on the (unidirectional) ring, so that data can be transmitted to all nodes in one cycle. Through geometric calculations for the ring, using delay values from Table 1, and incorporating latching delays (using ITRS data [47]) we obtained an operating frequency of greater than 2 GHz for the different sizes of CMPs we considered. Thus the photonic ring (and the communication network) was safely clocked at 2 GHz.

The power consumed in our hybrid photonic NoC can be divided into two parts: the power consumed in the electrical network and the power consumed in the optical ring. The static and dynamic power consumption of electrical routers in this work is based on results obtained from the Orion simulator [45], while the power consumption on optimally repeated and sized wires is obtained from the methodology in [46]. For calculating power consumption of the modulator driver and TIA power we used ITRS device projections [47] and standard circuit procedures. The average power consumption of each transmitter and receiver is 18.4 mW and 0.3 mW respectively, as derived from [24]. The power consumed by the transmitter dominates power consumed by the receiver because the size as well as the capacitance of the modulator is large, requiring a large driving circuit. The power consumed by the optical waveguide is almost independent of interconnect length, since the length is relatively short and consequently the optical power loss in the waveguide is negligible.

Table 1. Delay of optical components [24][44]

Technology	65 nm
Modulator driver (ps)	45.8
Modulator (ps)	52.1
Waveguide (ps/mm)	15.4
Photo Detector (ps)	0.5
Receiver	16.9

4.2 Results

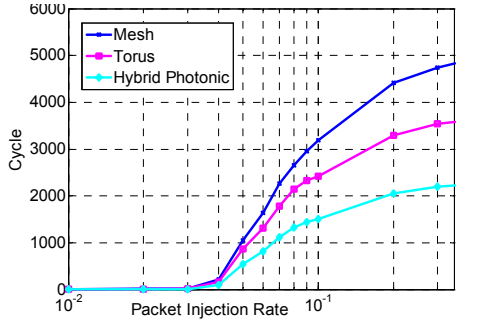
Photonic waveguides provide faster signal propagation compared to electrical interconnects because they do not suffer from RLC impedances. But in order to exploit the propagation speed advantage of photonic interconnects, electrical signals must be converted into light and then back into an electrical signal. This process requires a performance and power overhead that must be taken into account for an accurate analysis. To explore the impact on using our hybrid photonic NoC architecture on communication performance and power consumption, we performed several experiments. Results were obtained using a uniform random traffic model implemented on a cycle accurate NoC simulator in SystemC [48] that was derived from the Nirgam [49] and Noxim [50] NoC simulators. We targeted a 65nm process technology, and assume a 400 mm^2 die area. A high level floorplanner [51] is used to determine core placement and link lengths. The experiments are described in the following subsections.

4.2.1 Comparison with electrical mesh/torus NoCs

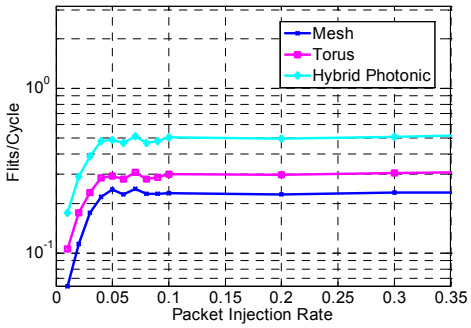
The first set of experiments compared our hybrid photonic NoC with traditional 2D all-electrical mesh and torus NoCs, for CMPs of varying complexity. Fig. 4(a)-(c) show the throughput, latency and power consumption for a 16 core (4×4) CMP architecture, with cores interconnected using the three NoC architecture alternatives. As packet injection rate from the cores increases, the latency, throughput, and power consumption of the all the NoC architectures initially increases. With further increase in packet injection rate, increased traffic congestion continues to cause an increase in average packet latency from the source to the

destination cores. Due to the photonic ring in our hybrid photonic NoC offloading a lot of the global communication from the electrical network, the congestion on the electrical network reduces, which results in lower average packet latency compared to the all-electrical mesh and torus architectures.

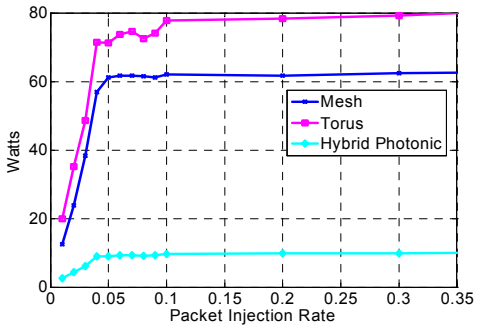
The addition of a high bandwidth photonic path also leads to a better average throughput response for our hybrid photonic NoC. This throughput however saturates quite rapidly as the injection rate leads to a greater load (and thus congestion) on all the NoC architectures. Finally, the rate of increase in power consumption of the NoC architectures also starts to saturate (after rapidly increasing initially) with increasing packet injection rate. Our hybrid photonic NoC can be seen to have a much lower power consumption compared to the all-electrical mesh and torus architectures. Results for a 36 core (6x6) CMP (not shown here for brevity) also demonstrate similar trends.



(a)

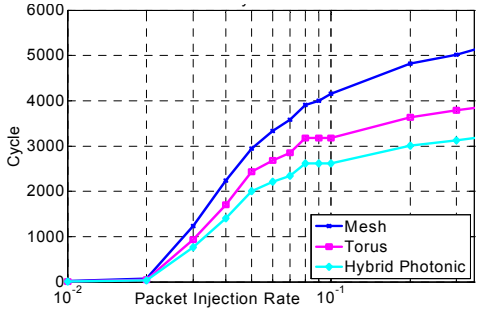


(b)

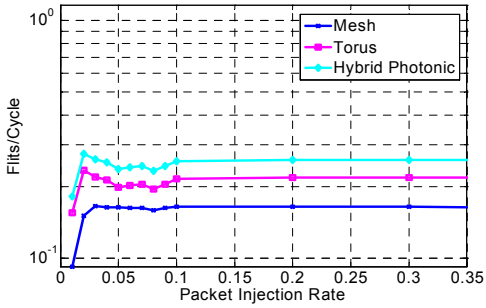


(c)

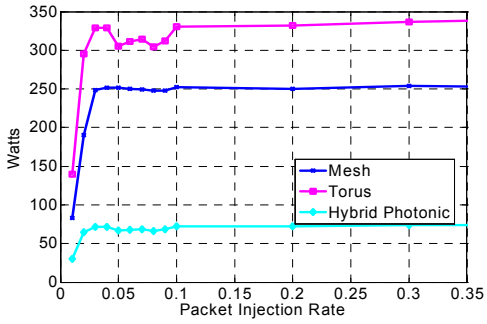
Figure 4. Hybrid photonic vs. electrical mesh vs. electrical torus for a 4x4 NoC (a) latency, (b) throughput, (c) power



(a)



(b)



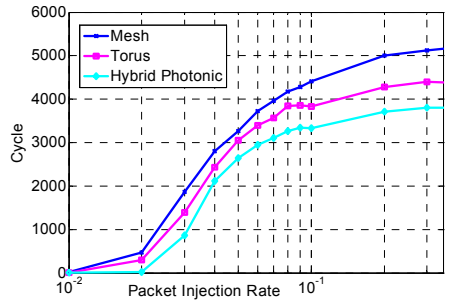
(c)

Figure 5. Hybrid photonic vs. electrical mesh vs. electrical torus for a 8x8 NoC (a) latency, (b) throughput, (c) power

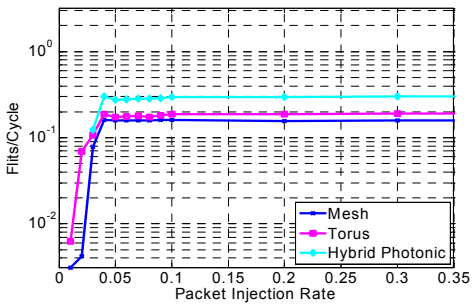
To analyze the impact of scaling application complexity on the performance improvement and power consumption savings of our hybrid photonic NoC, we repeated the above experiments for CMPs with 64 cores (8×8), as shown in Fig. 5(a)-(c) and 100 cores (10x10), as shown in Fig. 6(a)-(c). It can be seen that the performance (both latency and throughput) as well as power consumption of our hybrid photonic NoC is significantly better compared to the electrical mesh and torus NoCs at these core complexities. The power consumption in particular is several orders of magnitude lower for our hybrid photonic NoC. However, our analysis of link loads after simulation indicated that the utilization of the photonic ring was fairly low for such large CMPs. One reason for this is that we limited the optical region of influence to one, for all the experiments. As a result, the relative percentage of cores utilizing the optical path for global communication reduces with increasing CMP size. In the next subsection, we explore the impact of varying the region of optical influence for our proposed hybrid photonic NoC.

4.2.2 Impact of varying regions of optical influence

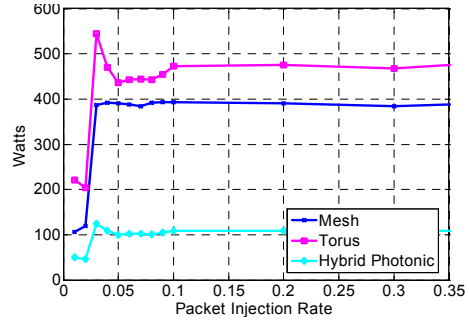
In order to improve the utilization of the photonic path especially for large sized CMPs, we explored varying the size of the regions of optical influence. Fig. 7 shows the % improvement in power consumption relative to the original case when the region of optical influence is increased from the original value of one, up to nine, for the 8×8 and 10×10 CMP sizes. It can be seen that there is a significant improvement with increasing optical influence, and the power consumption of the hybrid photonic NoC goes down. Note that as the region of influence increases, packets will also increasingly flow through potentially several nodes in the electrical portion of the NoC, before they can utilize the photonic path. Fig. 8 shows the performance impact of increasing the region of influence. Interestingly, there is a decrease in average latency and improvement in throughput as more and more packets utilize the photonic ring. Thus, increasing the region of optical influence allows us to reduce power consumption and also improve performance, making it an indispensible optimization for our proposed hybrid photonic NoC architecture.



(a)



(b)



(c)

Figure 6. Hybrid photonic vs. electrical mesh vs. electrical torus for a 10x10 NoC (a) latency, (b) throughput, (c) power

4.2.3 Impact of photonic serialization

In order to further reduce power consumption, we explore using data serialization for transfers over the photonic waveguide. The goal is to minimize the electro-optic circuitry and buffer sizes as well as switching activity, in order to reduce power consumption. Fig. 9 shows the reduction in power consumption for our hybrid photonic NoC as the degree of serialization is changed from the original un-serialized case (1×) to 2 (2:1 serialization) and 4 (4:1 serialization) for the 8×8 CMP case. From the figure it is clear that serialization has a notable impact on reducing power consumption, due to a reduction in the communication resources and switching activity (even after considering the power consumption overhead of the serializer/de-serializer circuitry). In addition to reducing power, serialization also reduces the number of photonic ring waveguides required, thus reducing the photonic layer complexity and cost.

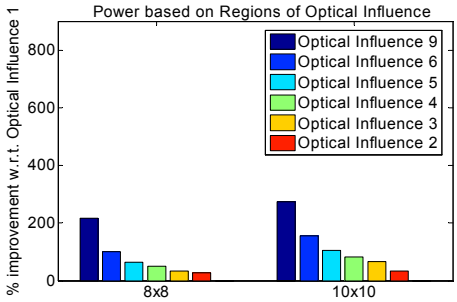
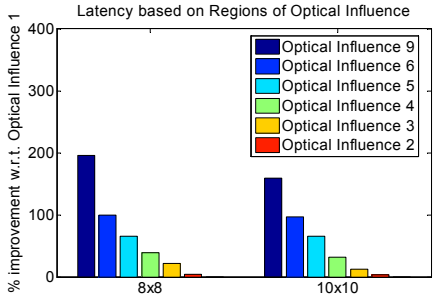
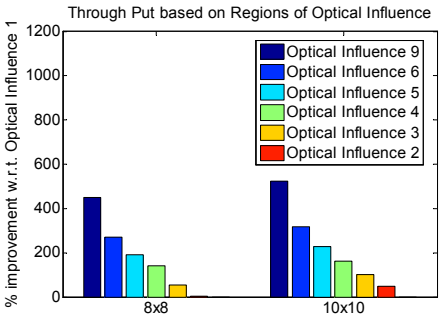


Figure 7. Reduction in power consumption with varying regions of optical influence



(a)



(b)

Figure 8. Relative impact of varying regions of optical influence on (a) average latency, (b) throughput

Serialization however entails a performance overhead since the number of bits transferred in a cycle gets reduced. Fig. 10 shows the reduction in throughput and an increase in latency as the degree

of serialization is increased, for the 8×8 CMP. It is clear that unlike the previous case where we optimized the region of optical influence, reducing power with serialization negatively impacts performance. Thus serialization must be used with great care. Serialization degrees of 2 and 4, as shown in the results here provide a good trade-off between power and performance. Higher degrees of serialization significantly reduce performance which is not practical, and thus those cases are not depicted in the results. Results for the 4×4 and 10×10 CMP cases showed a similar trend when subjected to serialization.

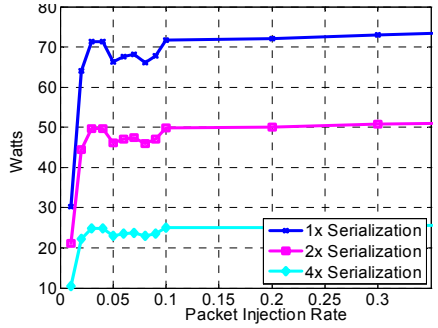
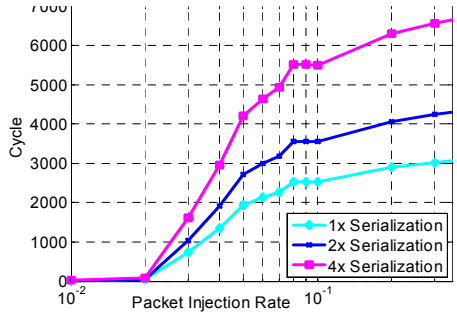
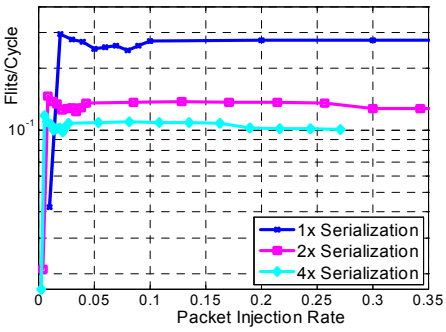


Figure 9. Impact of serialization on power consumption



(a)



(b)

Figure 10. Impact of serialization on (a) average latency, (b) throughput

4.2.3 Results Summary

Fig. 11 summarizes the performance and power comparison between our proposed hybrid photonic NoC and a traditional 2D all-electrical mesh NoC architecture, for 64 core (8×8) and 100 core (10×10) CMPs, with a packet injection rate of 0.35. Our hybrid photonic NoC architecture is configured with a serialization degree of 4 and a region of optical influence of 9. It can be seen from the figure that our carefully configured proposed architecture provides up to a 13× power reduction on average, an improvement of 1.9× for throughput and a reduction by 1.55× for latency compared to all-electrical mesh NoCs. Further improvements in performance can be obtained by varying the serialization degree, which trades off power consumption with performance. Ultimately, these results indicate a strong motivation to consider hybrid photonic NoCs for providing a high performance per watt communication infrastructure in future CMP applications.

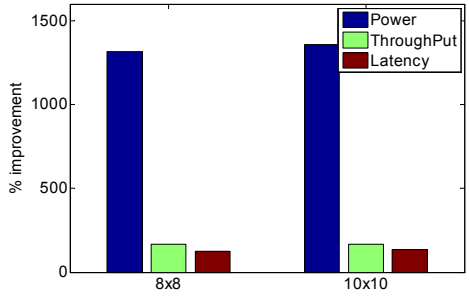


Figure 11. Overall improvement with hybrid photonic NoC

5. CONCLUSION

Reducing power consumption is essential to conserve battery life for mobile applications, and reduce the power expenditure of non-mobile applications. Future CMP applications with hundreds of cores will require a scalable communication fabric that can enable high performance per watt. It is not clear whether current 2D electrical NoCs can satisfy the performance requirements for future CMP applications with a highly constrained power budget. To address this challenge, in this work we propose a hybrid photonic NoC that utilizes a photonic ring on a dedicated silicon layer to complement a traditional electrical 2D mesh NoC. Results from our experiments show that the proposed architecture can result in a significant 13× reduction in power consumption, in addition to improvements in supported throughput and latency, compared to traditional 2D all-electrical mesh NoCs. The encouraging results from this work highlight the potential of using photonics on chip to meet the challenges of rising CMP complexity in the future.

6. REFERENCES

[1] D. Pham et al., "The Design and Implementation of a First-Generation CELL Processor," Proc. IEEE ISSCC, Feb. 2005.

[2] S. Vangal et al., "An 80-Tile 1.28 TFLOPS Network-on-Chip in 65 nm CMOS," Proc. IEEE ISSCC, 2007.

[3] Tilera Corporation. TILE64™ Processor. Product Brief. 2007.

[4] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Millberg, and D. Lindqvist, "Network on Chip: An Architecture for Billion Transistor Era," Proc. NorChip 2000.

[5] S. Pasricha, N. Dutt. "On-Chip Communication Architectures", Morgan Kauffman, Apr 2008.

[6] J.D. Owens, W.J. Dally, R. Ho, D.J. Jayasimha, S.W. Keckler, and L.-S. Peh, "Research Challenges for On-Chip Interconnection Networks," IEEE Micro, vol. 27, no. 5, pp. 96-108, Sept./Oct. 2007.

[7] Y. Hoskote et al. "A 5-GHz mesh interconnect for a teraflops processor" IEEE Micro, 27(5):51–61, Sept.-Oct. 2007.

[8] J.D. Owens et al. Research challenges for on-chip interconnection networks. IEEE Micro, 27(5):96–108, Sept.-Oct. 2007.

[9] S. Pasricha, N. Dutt, "Trends in Emerging On-Chip Interconnect Technologies", Proc. IPSJ Transactions on System LSI Design Methodology, Vol. 1, Sep 2008.

[10] A. Shacham, K. Bergman, and L.P. Carloni, "Photonic Networks-on Chip for Future Generations of Chip Multiprocessors", IEEE Trans. Computers, 57:9, Sep 2008.

[11] D. A. Miller, "Rationale and challenges for optical interconnects to electronic chips" Proc. IEEE, Jun 2000.

[12] R. Ramaswami and K.N. Sivarajan, Optical Networks: A Practical Perspective, second ed. Morgan Kaufmann, 2002.

[13] F. Xia, M.J. Rooks, L. Sekaric, and Y.A. Vlasov, "Ultra-Compact High Order Ring Resonator Filters Using Submicron Silicon Photonic Wires for On-Chip Optical Interconnects," Optics Express, 15(19) pp. 11934-11941, Sept. 2007.

[14] W.M.J. Green, M.J. Rooks, L. Sekaric, and Y.A. Vlasov, "Ultra-Compact, Low RF Power, 10 Gb/s Silicon Mach-Zehnder Modulator," Optics Express, vol. 15, no. 25, pp. 17106-17113, Dec. 2007.

[15] C.L. Schow, F. Doany, O. Liboiron-Ladouceur, C. Baks, D.M. Kuchta, L. Schares, R. John, and J.A. Kash, "160-Gb/s, 16-Channel Full-Duplex, Single-Chip CMOS Optical Transceiver," Proc. Optical Fiber Comm. Conf., paper OThG4, Mar. 2007.

[16] A. Biberman, B.G. Lee, K. Bergman, P. Dong, and M. Lipson, "Demonstration of All-Optical Multi-Wavelength Message Routing for Silicon Photonic Networks," Proc. Optical Fiber Comm. Conf., paper OTuF6, Mar. 2008.

[17] Y.A. Vlasov, W.M.J. Green, and F. Xia, "High-Throughput Silicon Nanophotonic Wavelength-Insensitive Switch for On-Chip Optical Networks," Nature Photonics, 2(4), Apr. 2008.

[18] W. Haensch, "Is 3D the Next Big Thing in Microprocessors?," Proc. International Solid State Circuits Conference (ISSCC) San Francisco, CA, 2007.

[19] J. W. Goodman et al., "Optical Interconnects for VLSI Systems," Proceedings of the IEEE, Vol. 72, No. 7, July 1984.

[20] J.H. Collet, F. Caignet, F. Sellaye, and D. Litaize, "Performance Constraints for Onchip Optical Interconnects," IEEE J. Selected Topics in Quantum Electronics, vol. 9, no. 2, pp. 425-432, Mar./Apr. 2003.

[21] M. Haurylau et al., "On-chip Optical Interconnect Roadmap: Challenges and Critical Directions," IEEE Journal of Selected Topics in Quantum Electronics, 12(6), Nov/Dec 2006.

[22] G. Tosik, et al., Power dissipation in optical and metallic clock distribution networks in new VLSI technologies, IEE Electronics Letters, vol. 40, no. 3, pp. 198-200, Feb. 2004.

[23] M.J. Kobrinsky et al., "On-Chip Optical Interconnects," Intel Technology J., vol. 8, no. 2, pp. 129-142, May 2004.

[24] G. Chen, H. Chen, M. Haurylau, N. Nelson, D. Albonesi, P. M. Fauchet, E. G. Friedman, "Predictions of CMOS Compatible On-Chip Optical Interconnect," Proc. SLIP, 2005.

[25] Ian O'Connor, "Optical solutions for system-level interconnect", Proc. SLIP, Feb 2004.

[26] A. M. Pappu A. B. Apsel, "Analysis of intrachip electrical and optical fanout", Applied Optics, 44(30), Oct 2005

[27] S. Pasricha, N. Dutt, "ORB: An On-chip Optical Ring Bus Communication Architecture for Multi-Processor Systems-on-Chip", IEEE/ACM ASPDAC 2008.

[28] M. Brie´re, et al., "System Level Assessment of an Optical NoC in an MPSoC Platform," Proc. DATE, Mar. 2007.

[29] D. Vantrease et al., "Corona: System Implications of Emerging Nanophotonic Technology", Proc. ISCA 2008.

[30] A. Shacham, K. Bergman, and L.P. Carloni, "The Case for Low- Power Photonic Networks on Chip," Proc. DAC, pp. 132-135, June 2007.

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

[32] C. Gunn, "CMOS Photonics for High-Speed Interconnects," IEEE Micro, vol. 26, no. 2, pp. 58-66, Mar./Apr. 2006.

[33] V.R. Almeida, C.A. Barrios, R.R. Panepucci, M. Lipson, M.A. Foster, D.G. Ouzounov, and A.L. Gaeta. All-optical switching on a silicon chip. Optics Letters, 29(24), Dec 2004.

[34] P. Rabiei, W. H. Steier, C. Zhang, and L. R. Dalton. Polymer micro-ring filters and modulators. Journal of Lightwave Technology, 20(11):1968–1975, Nov. 2002.

[35] S. Kawanishi, H. Takara, K. Uchiyama, I. Shake, and K. Mori. 3 Tbit/s (160 Gbit/s×19 channel) optical TDM and WDM transmission experiment. Electronic Letters, 35(10):826–827, 13 May 1999.

[36] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson. 12.5 Gbit/s carrier-injection-based silicon microring silicon modulators. Optics Express, 15(2):430–436, 22 Jan. 2007.

[37] A. Gupta, S. P. Levitan, L. Selavo, and D. M. Chiarulli. High-speed optoelectronics receivers in SiGe. In 17th Intl. Conf. on VLSI Design, pages 957–960, Jan. 2004.

[38] R. Dobkin, et al., "Parallel vs. Serial On-Chip Communication", Proc. SLIP 2008.

[39] A. Morgenshtein et al., "Comparative Analysis of Serial vs Parallel Links In NoC", Proc. SSOC, 2004.

[40] M. Ghoneima et al., "Serial-Link Bus: A Low-Power On-Chip Bus Architecture", Proc. ICCAD 2005.

[41] S. Kimura et al., "An On-Chip High Speed Serial Communication Method Based on Independent Ring Oscillators", Proc. ISSCC 2003.

[42] I-Chyn Wey et al., "A 2Gb/s High-Speed Scalable Shift-Register Based On-Chip Serial Communication Design for SoC Applications", Proc. ISCAS 2005.

[43] M. Saneei, A. Afzali-Kusha1, M. Pedram, "Two High Performance and Low Power Serial Communication Interfaces for On-chip Interconnects", Proc. CJECE 2008.

[44] I.-W. Hsieh, et al., "Ultrafast-Pulse Self-Phase Modulation and Third-Order Dispersion in Si Photonic Wire Wave-guides," Optics Express, 14(25), pp. 12380-12387, Dec. 2006.

[45] H.-S. Wang et al. Orion: A power-performance simulator for interconnection networks. In IEEE/ACM MICRO, Nov. 2002.

[46] R. Ho. On-Chip Wires: Scaling and Efficiency. Ph.D. dissertation, Dept. of Electrical Engineering, Stanford University, August 2003.

[47] The ITRS Technology Working Groups, http://public.itrs.net. International Technology Roadmap for Semiconductors (ITRS) 2007 Edition.

[48] SystemC initiative. www.systemc.org.

[49] Noxim simulator http://noxim.sourceforge.net/

[50] Nirgam simulator http://nirgam.ecs.soton.ac.uk/

[51] S. N. Adya, I. L. Markov, "Fixed-outline Floorplanning: Enabling Hierarchical Design", In IEEE Trans TVLSI, Dec. 2003.

[52] S. Bahirat, S. Pasricha, "Hybrid Photonic Interconnects for Chip Multiprocessors", CSU Technical Report 08-09.

[53] S. Pasricha, "Exploring Serial Vertical Interconnects for 3D ICs", IEEE/ACM DAC, Jul 2009