Silicon-Photonic Clos Networks for Global On-Chip Communication

Ajay Joshi,* Christopher Batten,* Yong-Jin Kwon,† Scott Beamer,† Imran Shamim*
Krste Asanović,† Vladimir Stojanović*

*Department of EECS, Massachusetts Institute of Technology, Cambridge, MA
†Department of EECS, University of California, Berkeley, CA

The on-chip communication network in manycore systems has become a critical component affecting system power, performance and programmer productivity. We explore the use of silicon-photonic technology to build low-diameter on-chip networks that scale well, and provide uniformly low latency and uniformly high bandwidth, which eases programming of these manycore systems. For this paper we consider a target system with 64 square tiles operating at 5 GHz on a 400 mm^2 chip in 22 nm technology.

1. On-Chip Network Design Space

We explore a spectrum of on-chip networks for manycore systems. At one end of the spectrum we have a mesh network that is attractive from a hardware design perspective due to short wires and low-radix routers. However, mapping a variety of applications, and in turn programming this network is difficult. At the other end, the strictly non-blocking crossbar network has long wires and high-radix router, which consume significant power and area even for communication between neighboring tiles. However, the crossbar properties – especially uniformly low latency, makes it attractive from a programming perspective.

It is therefore imperative to explore networks in the middle of the spectrum that could potentially provide uniformly low power and uniformly high performance, and at the same time ease the programming of these manycore systems. For example, researchers have proposed using concentrated mesh networks (cmesh) [1] to reduce hop count and in turn reduce the latency distribution, which makes programming easier. The cmesh topology can achieve similar throughput as a standard mesh with half the latency at the cost of longer channels and higher-radix routers. Similarly, a Clos [2] network that provides performance comparable to a crossbar but uses routers and point-to-point channels that are cheaper in terms of area and power can be used. However, unlike crossbar, a message in a Clos network needs multiple hops to reach its destination. In addition, a good routing algorithm is required to take advantage of path diversity and avoid network congestion.

We explore silicon photonics as a promising new interconnect technology which offers lower power, higher bandwidth density, and shorter latencies over long distances making them particularly effective for global point-to-point channels in low-diameter networks like crossbar, Clos etc. There have been other efforts in designing on-chip networks using silicon photonics – crossbar [4], mesh [3], etc. We present the photonic implementation of Clos as it provides better power–performance tradeoff compared to other networks. Figure 1 illustrates the various components in a typical wavelength-division multiplexed (WDM) photonic link used for on-chip communication. Light from an off-chip two-wavelength (λ_1, λ_2) laser source is carried by an optical fiber and then coupled into an on-chip waveguide. The waveguide carries the light past a series of transmitters, each using a resonant ring modulator to imprint the data on the corresponding wavelength. Modulated light continues through the waveguide to the other side of the chip where each of the two receivers use a tuned resonant ring filter to "drop" the corresponding wavelength from the waveguide into a local photodetector. The photodetector turns absorbed light into current, which is sensed by the electrical receiver. As the light waves travel through/along the various optical devices they experience optical losses, which directly affects system design as these losses set the required laser power. In addition to laser power, power is dissipated in thermal-tuning circuits that are used to maintain the resonance of ring filters and modulators under on-die temperature variations. In fact, for the global crossbar the power dissipated in thermally tuning the rings can be as large as 10 W, making it critical to explore thermally-insensitive photonic devices.

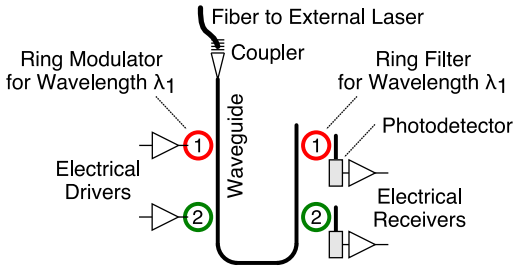
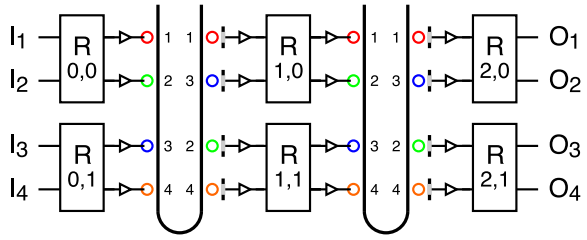


Figure 1: Photonic Components – Two point-to-point photonic links implemented with WDM.

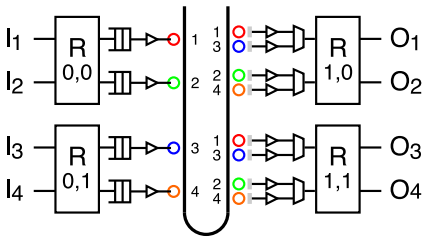
2. Evaluation

In this section, we use a detailed cycle-accurate simulator to study the performance and power of various electrical and photonic networks for the target system. We use synthetic traffic patterns based on a partitioned application model. Our exploration includes three electrical networks: a 2D mesh (*emesh*), two parallel cmesh networks with a concentration factor of four (*ecmeshX2*), and an 8-ary 3-stage Clos (*eclos*). Out of these three networks, the *ecmeshX2* network provided the best power-performance tradeoff. We also investigate two photonic networks: the distributed global crossbar (*pxbar*) and 8-ary 3-stage Clos network (*pclos*). In the crossbar, all global channels are implemented using photonic technology, while Figure 3 shows two possible implementations of the photonic Clos network. We use the implementation in Figure 3(a) here. Out of the two networks, the *pclos* provides better power-performance tradeoff. Here, we provide a comparison between the *ecmeshX2* and *pclos* networks.

Figure 2 (top row) shows the latency versus offered band-



(a) Clos with Photonic Point-to-Point Channels



(b) Clos with Photonic Middle Routers

Figure 3: Photonic implementation of 4-input 4-output 2-ary 3-stage Clos Networks with 6 2x2 routers– (a) four point-to-point photonic channels use WDM on each U-shaped waveguide. (b) the two middle routers ($R_{1,0-1}$) are implemented with photonic 2×2 crossbars on a single U-shaped waveguide. Number next to each ring indicates resonant wavelength.

width plots for *ecmeshX2* and *clos* network for various traffic patterns. To the first order, these plots are independent of the underlying technology. In Figure 2(a) we can see a wide distribution of the latency and saturation throughput for the four different traffic patterns due to the way applications get mapped to a *cmeshX2* network, and the use of dimension-ordered routing protocol. On the other hand, in case of the *clos* network (Figure 2(b)) there is a narrow distribution of latency and saturation throughput as the routing protocol necessitates the message to be always transported to an intermediate node and then to the destination node, and availability of path diversity.

Figure 2 (bottom row) shows the power dissipation versus offered bandwidth plots for *ecmeshX2* and *clos* network. For the *clos* network we have plotted power dissipation of both electrical and photonic design for comparison. Here while calculating power, we include switching power (50 fJ/bt @ 10 Gbps for photonic interconnects and 50 fJ/bt/mm @ 5 GHz clock for electrical interconnects) and leakage power in the electrical circuits and the thermal tuning power (1 μW/ring/K) for the optical devices. We do not include the off-chip laser power (≈3.3 W). From Figure 2(a) and 2(b), we can conclude that for global traffic patterns (UR, P2D and P8D), *pclos* consumes 2–3x lower power at comparable offered bandwidth. On the other hand, for local traffic patterns (P8C), *pclos* consumes comparable power at comparable offered bandwidth. The saturation throughput of the *pclos* for local traffic patterns can be improved by increasing the channel width (Figure 2(c)). Here too, compared to *ecmeshX2*, the *pclos* consumes lower (2–3x) and comparable power for global and local traffic patterns, respectively, at comparable throughput.

Acknowledgments

We would like to thank the various device groups at MIT for helping us in the design of photonic devices. This work was supported in part by Intel Corp. and DARPA awards W911NF-08-1-0134 and W911NF-08-1-0139.

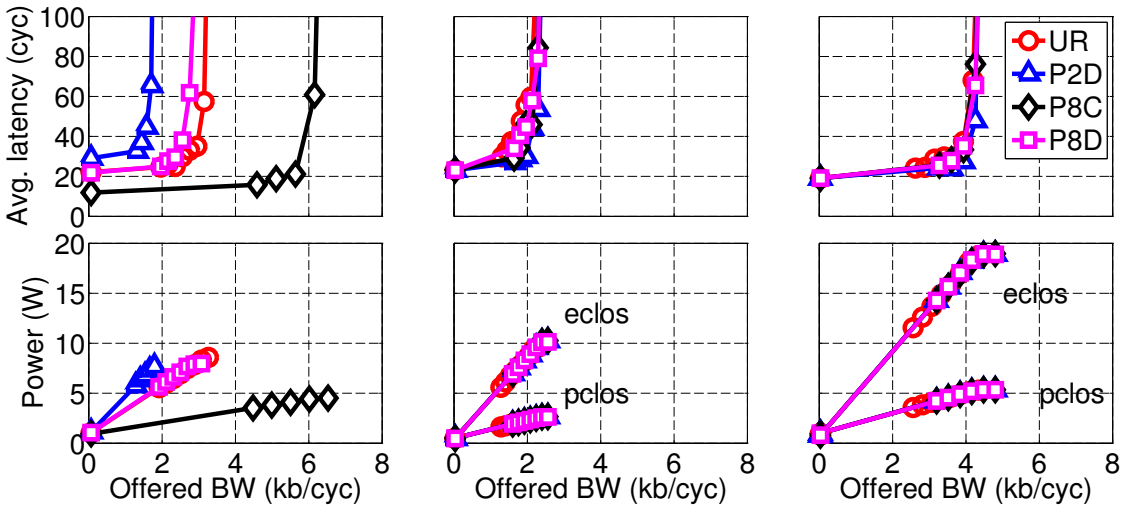
References

[1] J. Balfour and W. J. Dally. Design tradeoffs for tiled CMP on-chip networks. *Int'l Conf. on Supercomputing*, 2006.

[2] C. Clos. A study of non-blocking switching networks. *Bell System Technical Journal*, 32:406–424, 1953.

[3] M. Petracca et al. Design exploration of optical interconnection networks for chip multiprocessors. *Symp. on High-Performance Interconnects*, Aug. 2008.

[4] D. Vantrease et al. Corona: System implications of emerging nanophotonic technology. *ISCA*, June 2008.



(a) *ecmeshX2* - CW = 128 bits (b) *clos* - CW = 64 bits (c) *clos* - CW = 128 bits

Figure 2: Top row: Latency vs. Offered Bandwidth, Bottom row: Power vs. Offered Bandwidth – UR: Uniform random, P2D: Two tiles per partition where tiles are physically distributed across the chip, P8C: Eight tiles per partition where tiles are physically co-located and P8D: Eight tiles per partition where tiles are physically distributed across the chip. CW = Channel width.