Nano-Photonic Networks-on-Chip for Future Chip Multiprocessors

Cheng Li, Paul V. Gratz, and Samuel Palermo

1 Introduction

Parallel architectures, such as single-chip multiprocessors (CMPs), have emerged to address power consumption and performance scaling issues in current and future VLSI process technology. Networks-on-chip (NoCs), have concurrently emerged to serve as a scalable alternative to traditional, bus-based interconnection between processor cores. Conventional NoCs in CMPs use wide, point-to-point electrical links to relay cache-lines between private mid-level and shared last-level processor caches [1]. Electrical on-chip interconnect, however, is severely limited by power, bandwidth and latency constraints due to high-frequency loss of electrical traces and crosstalk from adjacent signals. These constraints are placing practical limits on the viability of future CMP scaling. For example, the efficiency of current state-of-the-art NoCs with simple CMOS inverter-based repeaters is near 2pJ/bit [2], allowing for only near 1TB/s throughput with a typical 20% allowance from the total 100W processor power budget. Power in electrical interconnects has been reported as high as 12.1W for a 48-core, 2D-mesh CMP at 2GHz [1], a significant fraction of the system's power budget. Furthermore, achieving application performance which scales with the number of cores requires extremely low latency communication to reduce the impact of serialization points within the code. However, communication latency in a typical NoC connected multiprocessor system increases rapidly as the number of nodes increases [3]. Worst-case, no-load communication latencies in a 64-node multi-core chip can reach as high as 50 cycles, nearly 1/2 the latency of an off-chip memory access. The communication requirements of future processing

Cheng Li
HP Laboratories, 1501 Page Mill Rd, Palo Alto, CA 94304, e-mail: cheng.li6@hp.com

Paul V. Gratz and Samuel Palermo
Texas A&M University, 3128 TAMU, College Station, TX 77843 e-mail: pgratz@gratz1.com;spalermo@ece.tamu.edu

systems makes traditional electrical on-chip networks prohibitive for future transformative extrascale computers.

Recently, monolithic silicon photonics have been proposed as a scalable alternative to meet future many-core systems bandwidth demands, by leveraging high-speed photonic devices [4, 5, 6], THz-bandwidth waveguides [7, 8], and immense bandwidth-density via wavelength-division-multiplexing (WDM) [9, 10]. Several NoC architectures leveraging the high bandwidth of silicon photonics have been proposed. These works can be categorized into two general types: 1). Hybrid optical/electrical interconnect architecture [11, 12, 13, 14], in which a photonic packet-switched network and an electronic circuit-switched control network are combined to respectively deliver large size data messages and short control messages; 2). Crossbar or Clos architectures, in which the interconnect is fully photonic [15, 16, 17, 18, 19, 20, 21, 22, 23]. Although these designs provide high and scalable bandwidth, they either suffer from relatively high latency due to the electrical control circuits for photonic path setup, or significant power/hardware overhead due to significant over-provisioned photonic channels. In future latency and power constrained CMPs, these characteristics will hobble the utility of photonic interconnect.

In this chapter, we propose LumiNOC [24], a novel PNoC architecture which addresses power and resource overheads due to channel over-provisioning, while reducing latency and maintaining high bandwidth in CMPs. LumiNoC utilizes integrated silicon waveguides that provide the potential to overcome electrical interconnect bottlenecks and greatly improve data transfer efficiency due to their flat channel loss over a wide frequency range and also relatively small crosstalk and electromagnetic noise [25]. By combining multiple data channels on a single waveguide via wavelength-division-multiplexing (WDM), LumiNoC greatly improves bandwidth density. Area-compact and energy-efficient silicon ring resonators are employed as the optical modulator and drop filter in the integrated WDM link. Silicon ring resonator modulators/filters offer advantages of small size, relative to Mach-Zehnder modulators [26], and increased filter functionality, relative to electro-absorption modulators [27]. The LumiNOC architecture makes three contributions: First, instead of conventional, globally distributed, photonic channels, requiring high laser power, we propose a novel channel sharing arrangement composed of sub-sets of cores in photonic subnets. Second, we propose a novel, purely photonic, distributed arbitration mechanism, dynamic channel scheduling, which achieves extremely low-latency without degrading throughput. Third, our photonic network architecture leverages the same wavelengths for channel arbitration and parallel data transmission, allowing efficient utilization of the photonic resources and lowering static power consumption. We show in a 64-node implementation that LumiNOC enjoys 50% lower latency at low loads and $\sim$40% higher throughput per Watt on synthetic traffic versus previous PNoCs. Furthermore, LumiNOC reduces latency $\sim$40% versus an electrical 2D mesh NoCs on PARSEC shared-memory, multithreaded benchmark workloads.

2 Silicon Photonic Devices

Fig. 1 shows a typical silicon photonics WDM link, where multiple wavelengths (λ_1-λ_4) generated by an off-chip continuous-wave (CW) laser are coupled into a silicon waveguide via an optical coupler. At transmit side, ring modulators insert data onto a specific wavelength through electro-optical modulation. These modulated optical signals propagate through the waveguide and arrive at the receiver side where ring filters drop the modulated optical signals of a specific wavelength at a receiver channel with photodetectors (PD) that convert the signals back to the electrical domain.

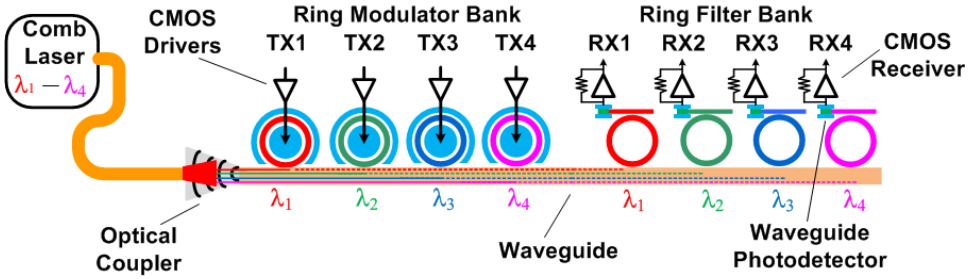


Fig. 1: Silicon ring resonator-based wavelength-division-multiplexing (WDM) link.

2.1 Laser Source

Laser source can either be a distributed feedback (DFB) laser bank [28], which consists of an array of DFB laser diodes, or a comb laser [29], which is able to generate multiple wavelengths simultaneously. Implementing a DFB laser bank for dense WDM (DWDM) photonic interconnects (e.g. 64 wavelengths) is quite challenging due to area and power budget constraints. This motivates a single broad-spectrum comb laser source, such as InAs/GaAs quantum dot comb lasers which can generate a large number of wavelengths in the 1100nm to 1320nm spectral range with typical channel spacing of 50-100GHz and optical power of 0.2-1mW per channel [29].

2.2 Microring Resonators (MRR)

MRRs can serve as either optical modulators for sending data or as filters for dropping and receiving data from an on-chip photonic network. A basic silicon ring modulator consists of a straight waveguide coupled with a circular waveguide with

diameters on the order of tens of micrometers, as shown in Fig. 2a. The two terminal device contains an input port, where the light source is coupled into, and a through port, where the modulated optical singal is coupled out. When the ring circumference equals an integer number of an optical wavelength, called the resonance condition, most of the input light is coupled into the circular waveguide and only a small amount of light can be observed at the through port. As a result, the through port spectrum displays a notch-shaped characteristic, shown in Fig. 2b. This resonance can be shifted by changing the effective refractive index of the waveguide through the free-carrier plasma dispersion effect [30] to implement the optical modulation. For example, the ring modulator exhibits low optical output power levels at the through port when the resonance is aligned well with the laser wavelength, while high optical power levels are displayed when the resonance shifts to a shorter wavelength (blue-shifts) due to the increase in the waveguide carrier density lowering the effective refractive index.

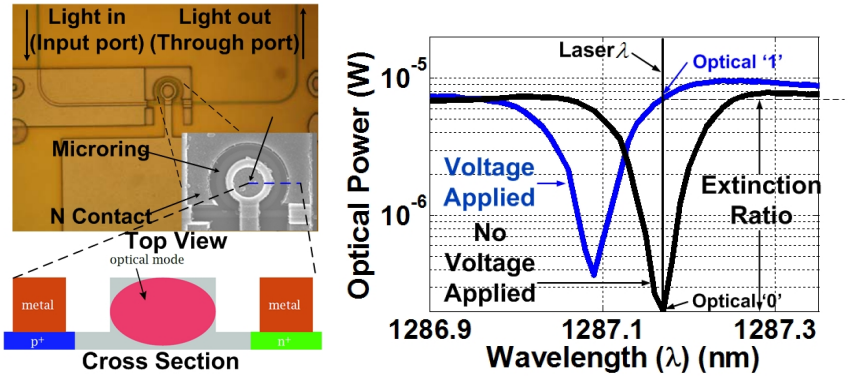
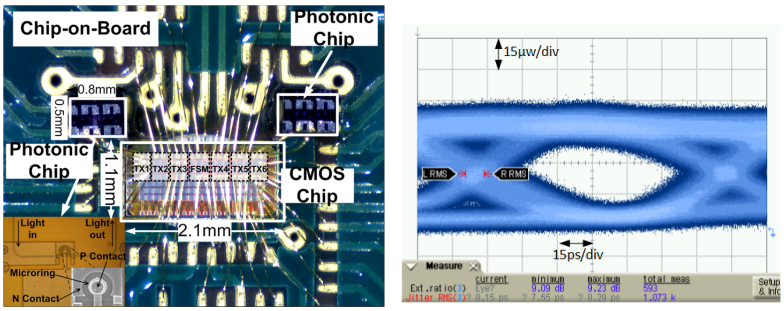


Fig. 2: (a) Top and cross section views of carrier-injection silicon ring resonator modulator, (b) optical spectrum at through port.

Two common implementations of silicon ring resonator modulators include carrier-injection devices [31], with an embedded p-i-n junction that is side-coupled with the circular waveguide and operating primarily in forward-bias, and carrier-depletion devices [32], with only a p-n junction side-coupled and operating primarily in reverse-bias. Although a depletion ring generally achieves higher modulation speeds relative to a carrier-injection ring due to the ability to rapidly sweep the carriers out of the junction, its modulation depth is limited due to the relatively low doping concentration in the waveguide to avoid excessive optical loss. In contrast, carrier-injection ring modulators can provide large refractive index changes and high modulation depths, but are limited by the relatively slow carrier dynamics

of forward-biased p-i-n junctions. Normally, this speed limitation can be alleviated with modulation and/or equalization techniques (e.g. pre-emphasis [33]).

An example of a carrier-injection ring modulator is the 5μm diameter device [34] shown in Fig. 3a, which was fabricated by HP Labs and exhibits a quality factor[1] of $\sim$9000. Here a chip-on-board test setup is utilized, with a 65nm CMOS driver [31] wire-bonded to silicon ring resonator chips for optical signal characterization. The measured optical eye diagram of this prototype is show in Fig. 3b. It achieves an extinction ratio[2] of 9.2dB at a modulation speed of 9Gb/s. The modulation efficiency is 500fJ/bit, including the electrical driver power. Adopting advanced CMOS processes (e.g. 16nm CMOS) and photonics integration techniques (e.g. flip-chip bonding or 3D integration) will further improve the optical modulation speed and energy efficiency. This provides strong motivation to leverage this photonic I/O architecture in a WDM system with multiple $\sim$10Gb/s channels on a single waveguide.



(a) (b)

Fig. 3: (a) Optical transmitter circuit prototype bonded for optical testing, (b) Measured ring modulator 9Gb/s optical eye diagram.

However, one important issue with MRR devices is their resonance wavelength's sensitivity to temperature variation, necessitating tuning to stablize the ring to resonate at the working wavelength. A commonly proposed resonance wavelength tuning technique is to adjust the device's temperature with a resistor implanted close to the photonic device to heat the waveguide, thus changing the refractive index [35, 36]. Thermal tuning efficiencies near 10-15μW/GHz have been demonstrated using approaches such as substrate removal and transfer for an SOI process [37] and deep-trench isolation for a bulk CMOS process [36]. Superior efficiencies in the 1.7-2.9μW/GHz have been achieved with localized substrate removal or undercutting [38, 39], but this comes at the cost of complex processing steps. One potential issue

[1] Quality factor characterizes a resonator's bandwidth relative to its center frequency. Higher Q indicates a lower rate of energy loss relative to the stored energy of the resonator.

[2] Extinction ratio is the ratio of two optical power levels of a modulated optical signal, expressed in dB.

with this approach is that the tuning speed, which is limited by the device thermal time constant ($\sim$ms), may necessitate long calibration times. Compared with the heater-based tuning approaches, a bias-based tuning method for carrier-injection rings has advantages of fast tuning speed and flexibility in the tuning direction, while displaying comparable tuning efficiency. A recent bias-based tuning scheme was reported with a power efficiency of 6.8μW/GHz, which includes the power of the tuning loop circuitry [31].

2.3 Silicon Waveguides

In photonic on-chip networks, silicon waveguides are used to carry the optical signals. In order to achieve higher aggregated bandwidth, multiple wavelengths are placed into a single waveguide in a wavelength-division-multiplexing (WDM) fashion. In this work, silicon nitride waveguides are assumed to be the primary transport strata. Similar to electrical wires, silicon nitride waveguides can be deployed into multiple strata to eliminate in-plane waveguide crossing, thus reducing the optical power loss [40].

2.4 Three-dimensional Integration

In order to optimize system performance and efficiently utilize the chip area, three-dimensional integration (3DI) is emerging for the integration of silicon nanophotonic devices with conventional CMOS electronics. In 3DI, the silicon photonic on-chip networks are fabricated into a separate silicon-on-insulator (SOI) die or layer with a thick layer of buried oxide (BOX) that acts as bottom cladding to prevent light leakage into the substrate. This photonic layer stacks above the electrical layers containing the compute tiles.

2.5 4-Tile Photonic Crossbar Example

Figure 4 shows a small CMP with 4 compute tiles interconnected by a fully connected crossbar PNoC. Each tile consists of a processor core, private caches, a fraction of the shared last-level cache, and a router connecting it to the photonic network. The photonic channel connecting the nodes is shown as being composed of MMRs (small circles), integrated photodetectors [6] and silicon waveguides [7, 8] (black lines connecting the circles). Transceivers (small triangles) mark the boundary between the electrical and photonic domain.

The simple crossbar architecture is implemented by provisioning four send channels, each utilizing the same wavelength in four waveguides, and four receive chan-

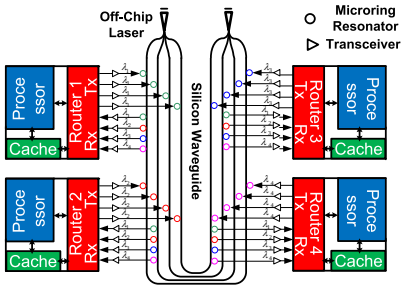


Fig. 4: Four-node fully connected photonic crossbar.

nels by monitoring four wavelengths in a single waveguide. Although this straightforward structure provides strictly non-blocking connectivity, it requires a large number of transceivers $O(r^2)$ and long waveguides crossing the chip, where r is the crossbar radix, thus this style of crossbar is not scalable to a significant number of nodes. Researchers have proposed a number of PNoC architectures more scalable than fully connected crossbars, as described below.

3 Photonic Network-on-Chip Architecture Survey

Many PNoC architectures have been proposed which may be broadly categorized into four basic architectures: 1) Electrical-photonic 2) Crossbar 3) Multi-stage and 4) Free-space designs.

Electrical-Photonic Designs: Shacham et al. propose a hybrid electrical-photonic NoC using electrical interconnect to coordinate and arbitrate a shared photonic medium [11, 12]. These designs achieve very high photonic link utilization by effectively trading increased latency for higher bandwidth. While increased bandwidth without regard for latency is useful for some applications, it eschews a primary benefit of PNoCs over electrical NoCs, low latency. Hendry et al. addressed this issue by introducing an all optical mesh network with photonic time division multiplexing (TDM) arbitration to set up communication path. However, the simulation results show that system still suffers from relatively high average latency [41].

Crossbar Designs: Other PNoC work attempts to address the latency issue by providing non-blocking point-to-point links between nodes. In particular, several works propose crossbar topologies to improve the latency of multi-core photonic interconnect. Fully connected crossbars [17] do not scale well, but researchers have examined channel sharing crossbar architectures, called Single-Write-Multiple-Read (SWMR) or Multiple-Write-Single-Read (MWSR), with various arbitration mechanisms for coordinating shared sending and/or receiving channels. Vantrease et al. proposed Corona, a MWSR crossbar, in which each node listens on the dedicated channel, but with the other nodes competing to send data on this channel [20, 21]. To implement arbitration at sender side, the author implemented a token channel

[21] or token slot [20] approach similar to token rings used in early LAN network implementations. Alternately, Pan et al. proposed Firefly, a SWMR crossbar design, with a dedicated sending channel for each node, but all the nodes in a crossbar listen on all the sending channels [19]. Pan et al. proposed broadcasting the flit-headers to specify a particular receiver.

In both SWMR and MWSR crossbar designs, over-provisioning of dedicated channels, either at the receiver (SWMR) or sender (MWSR), is required, leading to under utilization of link bandwidth and poor power efficiency. Pan et al. also proposed a channel sharing architecture, FlexiShare [18], to improve the channel utilization and reduce channel over-provisioning. The reduced number of channels, however, limit the system throughput. In addition, FlexiShare requires separated dedicated arbitration channels for sender and receiver sides, incurring additional power and hardware overhead.

Two designs propose to manage laser power consumption at runtime. Chen and Joshi propose to switch off portions of the network based on the bandwidth requirements [42]. Zhou and Kodi propose a method to predict future bandwidth needs and scale laser power appropriately [43].

Multi-stage Designs: Joshi et al. proposed a photonic multi-stage Clos network with the motivation of reducing the photonic ring count, thus reducing the power for thermal ring trimming [15]. Their design explores the use of a photonic network as a replacement for the middle stage of a three-stage Clos network. While this design achieves an efficient utilization of the photonic channels, it incurs substantial latency due to the multi-stage design.

Koka et al. present an architecture consisting of a grid of nodes where all nodes in each row or column are fully connected by a crossbar [22]. To maintain full-connectivity of the network, electrical routers are used to switch packets between rows and columns. In this design, photonic "grids" are very limited in size to maintain power efficiency, since fully connected crossbars grow at $O(n^2)$ for the number of nodes connected. Kodi and Morris propose a 2-D mesh of optical MWSR crossbars to connect nodes in the x and y dimensions [44]. In a follow-on work by the same authors Morris et al. [45] proposed a hybrid multi-stage design, in which grid rows (x-dir) are subnets fully connected with a photonic crossbar, but different rows (y-dir) are connected by a token-ring arbitrated shared photonic link. Bahirat and Pasricha propose an adaptable hybrid design in which a 2-D Mesh electrical network is overlaid with a set of photonic rings [46].

Free-Space Designs: Xue et al. present a novel free-space optical interconnect for CMPs, in which optical free-space signals are bounced off of mirrors encapsulated in the chip's packaging [47]. To avoid conflicts and contention, this design uses in-band arbitration combined with an acknowledgment based collision detection protocol.

4 Power Efficiency in PNoCs

Power efficiency is an important motivation for photonic on-chip interconnect. In photonic interconnect, however, the static power consumption (due to off-chip laser, ring thermal tuning, etc) dominates the overall power consumption, potentially leading to energy-inefficient photonic interconnects. In this section, prior photonic NoCs are examined in terms of static power efficiency. Bandwidth per watt is used as the metric to evaluate power efficiency of photonic interconnect architectures, showing that it can be improved by optimizing the interconnect topology, arbitration scheme and photonic device layout.

Channel Allocation: We first examine channel allocation in prior photonic interconnect designs. Several previous photonic NoC designs, from fully connected crossbars [17] to the blocking crossbar designs [16, 18, 19, 20, 21], provide extra channels to facilitate safe arbitration between sender and receiver. Although conventional photonic crossbars achieve nearly uniform latency and high bandwidth, channels are dedicated to each node and cannot be flexibly shared by the others. Due to the unbalanced traffic distribution in realistic workloads [48], channel bandwidth cannot be fully utilized. This leads to inefficient energy usage, since the static power is constant regardless of traffic load. Over-provisioned channels also implies higher ring resonator counts, which must be maintained at the appropriate trimming temperature, consuming on-chip power. Additionally, as the network size increases, the number of channels required may increase quadratically, complicating the waveguide layout and leading to extra optical loss. An efficient photonic interconnect must solve the problem of efficient channel allocation. Our approach leverages this observation to achieve lower power consumption than previous designs.

Topology and Layout: Topology and photonic device layout can also cause unnecessary optical loss in the photonic link, which in turn leads to greater laser power consumption. Many photonic NoCs globally route waveguides in a bundle, connecting all the tiles in the CMP [16, 19, 20, 21]. In these designs, due to the unidirectional propagation property of optical transmission, the waveguide must be routed to reach each node twice (double-back), such that the signal being modulated by senders on the outbound path may be received by all possible receivers. The length of these double-back waveguides leads to significant laser power losses over the long distance.

Figure 5 shows the optical link budgets for the photonic data channel of Corona [21], Firefly [19], Clos [15] and LumiNOC under same radix and chip area, based on our power model (described in Section 6.5). Flexishare [18] is not compared, since not enough information was provided in the paper to estimate the optical power budget at each wavelength. The figure shows that waveguide losses dominate power loss in all three designs. This is due to the long waveguides required to globally route all the tiles on a chip. For example, the waveguide length in Firefly and Clos network in a 400 mm^2 chip are estimated to be 9.5cm and 5.5cm, respectively. This corresponds to 9.5dB and 5.5dB loss in optical power, assuming the waveguide loss is 1dB/cm [15]. Moreover, globally connected tiles imply a relatively higher number of rings on each waveguide, leading to higher ring through loss. Despite a single--

run, bi-directional architecture, even the Clos design shows waveguide loss as the largest single component.

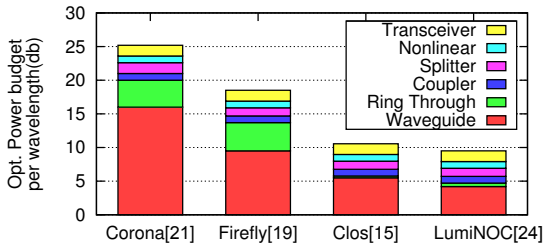


Fig. 5: Optical link budgets for the photonic data channels of various photonic NoCs.

In contrast to other losses (e.g. coupler and splitter loss, filter drop loss and photodetector loss) which are relatively independent of interconnect architecture, waveguide and ring through loss can be reduced through layout and topology optimization. We propose a network architecture which reduces optical loss by decreasing individual waveguide length as well as the number of rings along the waveguide.

Arbitration Mechanism: The power and overhead introduced by the separated arbitration channels or networks in previous photonic NoCs can lead to further power efficiency losses. Corona, a MWSR crossbar design, requires a token channel or token slot arbitration at sender side [20, 21]. Alternatively, Firefly [19], a SWMR crossbar design, requires head-flit broadcasting for arbitration at receiver side, which is highly inefficient in PNoCs. FlexiShare [18] requires both token stream arbitration and head-flit broadcast. These arbitration mechanisms require significant overhead in the form of dedicated channels and photonic resources, consuming extra optical laser power. For example, the radix-32 Flexishare [18] with 16 channels requires 416 extra wavelengths for arbitration, which accounts for 16% of the total wavelengths in addition to higher optical power for a multi-receiver broadcast of head-flits. Arbitration mechanisms are a major overhead for these architectures, particularly as network radix scales.

There is a clear need for a PNoC architecture that is energy-efficient and scalable while maintaining low latency and high bandwidth. In the following sections, we propose the LumiNOC architecture which reduces the optical loss by partitioning the global network into multiple smaller sub-networks. Furthermore, the proposed novel arbitration scheme leverages the same wavelengths for channel arbitration and parallel data transmission to efficiently utilize the channel bandwidth and photonic resources, without dedicated arbitration channels or networks which lower efficiency or add power overhead to the system.

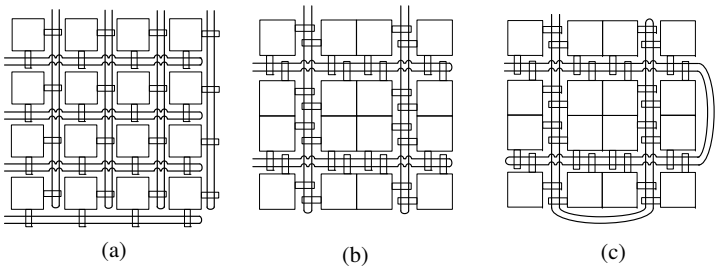


Fig. 6: LumiNOC interconnection of CMP with 16 tiles - (a) One- (b) Two- and (c) Four-rows interconnection.

5 LumiNOC Architecture

In our analysis of prior PNoC designs, we have found that a significant amount of laser power consumption was due to the waveguide length required for propagation of the photonic signal across the entire network. Based on this, the LumiNOC design breaks the network into several smaller networks (subnets), with shorter waveguides. Figure 6 shows three example variants of the LumiNOC architecture with different subnet sizes, in an example 16-node CMP system: the one-row, two-rows and four-rows designs (note: 16-nodes are shown to simplify explanation, in Section 6 we evaluate a 64-node design). In the one-row design, a subnet of four tiles is interconnected by a photonic waveguide in the horizontal orientation. Thus four non-overlapping subnets are needed for the horizontal interconnection. Similarly four subnets are required to vertically interconnect the 16 tiles. In the two-row design, a single subnet connects 8 tiles while in the four-row design a single subnet touches all 16 tiles. In general, all tiles are interconnected by two different subnets, one horizontal and one vertical. If a sender and receiver do not reside in the same subnet, transmission requires a hop through an intermediate node's electrical router. In this case, transmission experiences longer delay due to the extra O/E-E/O conversions and router latency. To remove the overheads of photonic waveguide crossings required by the orthogonal set of horizontal and vertical subnets, the waveguides can be deposited into two layers with orthogonal routing [40].

Another observation from prior photonic NoC designs is that channel sharing and arbitration have a large impact on design power efficiency. Efficient utilization of the photonic resources, such as wavelengths and ring resonators, is required to yield the best overall power efficiency. To this end, we leverage the same wavelengths in the waveguide for channel arbitration and parallel data transmission, avoiding the power and hardware overhead due to the separated arbitration channels or networks. Unlike the over-provisioned channels in conventional crossbar architectures, channel utilization in LumiNOC is improved by multiple tiles sharing a photonic channel.

A final observation from our analysis of prior photonic NoC design is that placing many wavelengths within each waveguide through deep wavelength-division

multiplexing (WDM) leads to high waveguide losses. This is because the number of rings that each individual wavelength encounters as it traverses the waveguide is proportional to the number of total wavelengths in the waveguide times the number of waveguide connected nodes, and each ring induces some photonic power losses. We propose to limit LumiNOC's waveguides to a few frequencies per waveguide and increase the count of waveguides per subnet, to improve power efficiency with no cost to latency or bandwidth, a technique we call "ring-splitting". Ring-splitting is ultimately limited by the tile size and optical power splitting loss. Assuming a reasonable waveguide pitch of 15μm required for layout of microrings which have a diameter of 5μm [31], this leaves 5μm clearance to avoid optical signal interference between two neighboring rows of rings.

5.1 LumiNOC Subnet Design

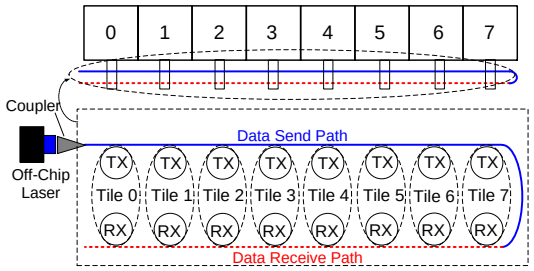


Fig. 7: One-row subnet of eight nodes. Circles (TX and RX) represent groups of rings; one dotted oval represents a tile.

Figure 7 details the shared channel for a LumiNOC one-row subnet design. Each tile contains W modulating "Tx rings" and W receiving "Rx Rings", where W is the number of wavelengths multiplexed in the waveguide. Since the optical signal unidirectionally propagates in the waveguide from its source at off-chip laser, each node's Tx rings are connected in series on the "Data Send Path", shown in a solid line from the laser, prior to connecting each node's Rx rings on the "Data Receive Path", shown in a dashed line. In this "double-back" waveguide layout, modulation by any node can be received by any other node; furthermore, the node which modulates the signal may also receive its own modulated signal, a feature that is leveraged in our collision detection scheme in the arbitration phase. The same wavelengths are leveraged for arbitration and parallel data transmission.

During data transmission, only a single sender is modulating on all wavelengths and only a single receiver is tuned to all wavelengths. However, during arbitration (*i.e.* any time data transfer is not actively occurring) the Rx rings in each node are tuned to a specific, non-overlapping set of wavelengths. Up to half of the wave-

lengths available in the channel are allocated to this arbitration procedure. with the other half available for credit packets as part of credit-based flow control. This particular channel division is designed to prevent optical broadcasting, the state when any single wavelength must drive more than one receiver, which if allowed would severely increase laser power [49]. Thus, at any given time a multi-wavelength channel with N nodes may be in one of three states: **Idle** - All wavelengths are unmodulated and the network is quiescent. **Arbitration** - One more sender nodes are modulating N copies of the arbitration flags; one copy to each node in the subnet (including itself) with the aim to gain control of the channel. **Data Transmission** - Once a particular sender has established ownership of the channel, it modulates all channel wavelengths in parallel with the data to be transmitted.

In the remainder of this section, we detail the following: *Arbitration* - the mechanism by which the photonic channel is granted to one sender, avoiding data corruption when multiple senders wish to transmit, including *Dynamic Channel Scheduling*, the means of sender conflict resolution, and *Data Transmission* - the mechanism by which data is transmitted from sender to receiver. *Credit Return* is also discussed.

Arbitration: We propose an optical collision detecting and dynamic channel scheduling technique to coordinate access of the shared photonic channel. This approach achieves efficient channel utilization without the latency of electrical arbitration schemes [11, 12], or the overhead of wavelengths and waveguides dedicated to standalone arbitration [21, 19, 18]. In this scheme, a sender works together with its own receiver to ensure message delivery in the presence of conflicts.

Receiver: Once any receiver detects an arbitration flag, it will take one of three actions: if the arbitration flag is uncorrupted (*i.e.* the sender flag has a 0 in only one location indicating single-sender) and the forthcoming message is destined for this receiver, it will enable all its Rx rings for the indicated duration of the message, capturing it. If the arbitration flags are uncorrupted, but the receiver is not the intended destination, it will detune all of its Rx rings for the indicated duration of the message to allow the recipient sole access. Finally, if a collision is detected, the receiver circuit will enter the **Dynamic Channel Scheduling** phase (described below).

Sender: To send a packet, a node first waits for any on-going messages to complete. Then, it modulates a copy of the arbitration flags to the appropriate arbitration wavelengths for each of the N nodes. The arbitration flags for an example 4-node subnet are depicted in Figure 8. The arbitration flags are a t_{arb} cycle long header (2 in this example) made up of the destination node address (D0-D1), a bimodal packet size indicator (Ln) for the two supported payload lengths (64-bit and 576-bit), and a "1-hot" encoded source address (S0-S3) (*i.e.* the source address is coded so that each valid encoding for a given source will have exactly one bit set) which serves as a guard band or collision detection mechanism: since the subnet is operated synchronously, any time multiple nodes send overlapping arbitration flags, the "1-hot" precondition is violated and all nodes are aware of the collision. We leverage self-reception of the arbitration flag to detect collision. Right after sending, the node monitors the incoming arbitration flags. If they are uncorrupted (*i.e.* only one bit is set in the source address), then the sender succeeded in arbitrating the channel and the two nodes proceed to the **Data Transmission** phase. If the arbitration flags

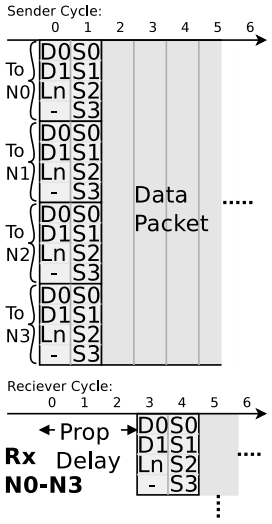


Fig. 8: Arbitration on a 4-node subnet.

are corrupted (*i.e.* more than one bit is set in the source address), then a conflict has occurred. Any data already sent is ignored and the conflicting senders enter the **Dynamic Channel Scheduling** regime (described below).

The physical length of the waveguide incurs a propagation delay, t_{pd} (cycles), on the arbitration flags traversing the subnet. The "1-hot" collision detection mechanism will only function if the signals from all senders are temporally aligned, so if nodes are physically further apart than the light will travel in 1 cycle, they will be in different clocking domains to keep the packet aligned as it passes the final sending node. Furthermore, the arbitration flags only start on cycles that are an integer multiple of the $t_{pd}+1$ to assure that no nodes started arbitration during the previous t_{slot} and that all possibly conflicting arbitration flags are aligned. This means that conflicts only occur on arbitration flags, not with data.

Note that a node will not know if it has successfully arbitrated the channel until after $t_{pd}+t_{arb}$ cycles, but will begin data transmission after t_{arb}. In the case of an uncontested link, the data will be captured by the receiver without delay. Upon conflict, senders cease sending (unusable) data.

As as an example, say that the packet in Figure 8 is destined for node 2 with no conflicts. At cycle 5, Nodes 1, 3, and 4 would detune their receivers, but node 2 would enable them all and begin receiving the data flits.

If the subnet size were increased without proportionally increasing the available wavelengths per subnet, then the arbitration flags will take longer to serialize as more bits will be required to encode the source and destination address. If, however, additional wavelengths are provisioned to maintain the bandwidth/node, then the additional arbitration bits are sent in parallel. Thus the general formula for $t_{arb} =$

$ceil(1+N+log_2(N)/\lambda)$ where N is the number of nodes and λ is the number of wavelengths per arbitration flag.

Dynamic Channel Scheduling: Upon sensing a conflicting source address, all nodes identify the conflicting senders and a dynamic, fair schedule for channel acquisition is determined using the sender node index and a global cycle count (synchronized at startup): senders transmit in $(n+cycle) \mod N$ order. Before sending data in turn, each sender transmits an abbreviated version of the arbitration flags: the destination address and the packet size. All nodes tune in to receive this, immediately followed by the **Data Transmission** phase with a single sender and receiver for the duration of the packet. Immediately after the first sender sends its last data flit, the next sender repeats this process, keeping the channel occupied until the last sender completes. After the dynamic schedule completes, the channel goes idle and any node may attempt a new arbitration to acquire the channel as previously described.

Data Transmission: In this phase the sender transmits the data over the photonic channel to the receiving node. All wavelengths in the waveguide are used for bit-wise parallel data transmission, so higher throughput is expected when more wavelengths are multiplexed into the waveguide. Two packet payload lengths, 64-bit for simple requests and coherence traffic and 576-bit for cache line transfer, are supported.

Credit Return: At the beginning of any arbitration phase (assuming the channel is not in use for Data Transmission), 1/2 of the wavelengths of the channel are reserved for credit return from the credit return transmitter (*i.e.* the router which has credit to return) to the credit return receiver (*i.e.* the node which originally sent the data packet and now must be notified of credit availability). Similar to the arbitration flags, the wavelengths are split into N different sub-channels, each one dedicated to a particular credit return receiver. Any router which has credit to send back may then modulate its credit return flag onto the sub-channel to the appropriate credit return receiver. The credit return flag is encoded similarly to the arbitration flag. In the event of a collision between two credit return senders returning credit to the same receiver, no retransmission is needed as the sender part of the flag will uniquely identify all nodes sending credit back to this particular credit return receiver. Credit is returned on a whole-packet basis, rather than a flit basis to decrease overheads. The packet size bit *Ln* is not used in the credit return flag; credit return receivers must keep a history of the packet sizes transmitted so that the appropriate amount of credit is returned.

5.2 Router Microarchitecture

The electrical router architecture for LumiNOC is shown in Figure 9. Each router serves both as an entry point to the network for a particular core, as well as an intermediate node interconnecting horizontal and vertical subnets. If a processor must send data to another node on the same vertical or horizontal subnet, packets

are switched from the electrical input port to the vertical photonic output port with one E/O conversion. Packets which are destined for a different subnet must be first routed to an intermediate node via the horizontal subnet before being routed on the vertical subnet. Each input port is assigned with a particular virtual-channel (VCs) to hold the incoming flits for a particular sending node. The local control unit performs routing computation, virtual-channel allocation and switching allocation in crossbar. The LumiNOC router's complexity is similar to that of an electrical, bi-directional, 1-D ring network router, with the addition of the E/O-O/E logic.

6 Evaluation

In this section, we describe a particular implementation of the LumiNOC architecture and analyze its performance and power efficiency.

6.1 64-Core LumiNOC Implementation

Here we develop a baseline physical implementation of the general LumiNOC architecture specified in Section 5 for the evaluation of LumiNOC against competing PNOC architectures. We assume a 400 mm^2 chip implemented in a 22nm CMOS process and containing 64 square tiles that operate at 5GHz, as shown in Figure 10. A 64-node LumiNOC design point is chosen here as a reasonable network size which could be implemented in a 22nm process technology. Each tile contains a processor core, private caches, a fraction of the shared last-level cache, and a router connecting it to one horizontal and one vertical photonic subnet. Each router input port contains seven virtual channels (VCs), each five flits deep. Credit based flow control is implemented via the remainder of the photonic spectrum not used for arbitration during arbitration periods in the network.

A 64-node LumiNOC may be organized into three different architectures: the one-row, two-row and four-row designs (shown in Figure 6), which represent a trade-off between interconnect power, system throughput and transmission latency. For example, power decreases as row number increases from one-row to two-row, since the single waveguide is roughly with the same length, but fewer waveguides are required. The low-load latency is also reduced due to more nodes residing in the same subnet, reducing the need for intermediate hops via an electrical router. The two-row subnet design, however, significantly reduces throughput due to the reduced number of transmission channels. As a result, we choose the "one-row" subnet architecture of Figure 6a, with 64-tiles arranged as shown in Figure 10 for the remainder of this section. In both the horizontal and vertical axes there are 8 subnets which are formed by 8 tiles that share a photonic channel, resulting in all tiles being redundantly interconnected by two subnets. Silicon nitride waveguides are assumed to be the primary transport strata. Similar to electrical wires, silicon nitride waveg-

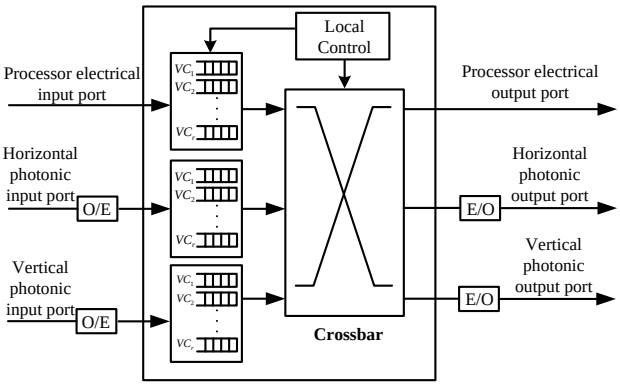


Fig. 9: Router microarchitecture.

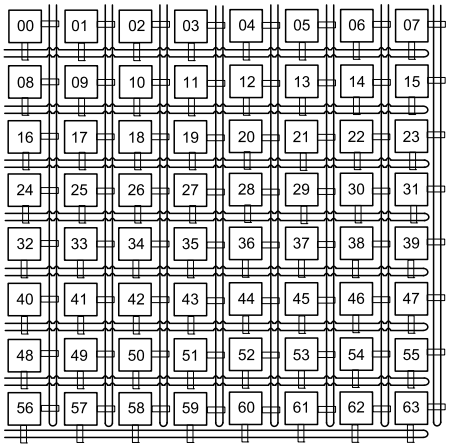


Fig. 10: One-row LumiNOC with 64 tiles.

uides can be deployed into multiple strata to eliminate in-plane waveguide crossing, thus reducing the optical power loss [40]. In order to optimize system performance and efficiently utilize the chip area, three-dimensional integration (3DI) is emerging for the integration of silicon nanophotonic devices with conventional CMOS electronics. In 3DI, the silicon photonic on-chip networks are fabricated into a separate silicon-on-insulator (SOI) die or layer with a thick layer of buried oxide (BOX) that acts as bottom cladding to prevent light leakage into the substrate. This photonic layer stacks above the electrical layers containing the compute tiles.

As a general trend, multirow designs tend to decrease power consumption in the router as fewer router hops are required to cover more of the network. Because of the diminishing returns in terms of throughput as channel width increases, however, congestion increases and the bandwidth efficiency drops. Further, the laser power

grows substantially for a chip as large as the one described here. For smaller floorplans, however, multi-row LumiNOC would be an interesting design point.

We assume a 10GHz network modulation rate, while the routers and cores are clocked at 5GHz. Muxes are placed on input and output registers such that on even network cycles, the photonic ports will interface with the lower half of a given flit and on odd, the upper half. With a 400 mm^2 chip, the effective waveguide length is 4.0 cm, yielding a propagation delay of $t_{pd} = 2.7$ 10GHz network cycles.

When sender and receiver reside in the same subnet, data transmission is accomplished with a single hop, *i.e.* without a stop in an intermediate electrical router. Two hops are required if sender and receiver reside in different subnets, resulting in a longer delay due to the extra O/E-E/O conversion and router latency. The "one-row" subnet based network implies that for any given node 15 of the 63 possible destinations reside within one hop, the remaining 48 destinations require two hops.

Link Width versus Packet Size: Considering the link width, or the number of wavelengths per logical subnet, if the number of wavelengths and thus channel width is increased, it should raise ideal throughput and theoretically reduce latency due to serialization delay. We are constrained, however, by the 2.7 network cycle propagation delay of the link (t_{pd} above), and the small packet size of single cache line transfers in typical CMPs. There is no advantage to sending the arbitration flags all at once in parallel when additional photonic channels are available; the existing bits would need to be replaced with more guard bits to provide collision detection. Thus, the arbitration flags would represent an increasing overhead. Alternately, if the link were narrower, the 2.7 cycle window would be too short to send all the arbitration bits and a node would waste time broadcasting arbitration bits to all nodes after it effectively "owns" the channel. Thus, the optimal link width is 64 wavelengths under our assumptions for clock frequency and waveguide length.

If additional spectrum or waveguides are available, then we propose to implement multiple parallel, independent **Network Layers**. Instead of one network with a 128-bit data path, there will be two parallel 64-bit networks. This allows us to exploit the optimal link width while still providing higher bandwidth. When a node injects into the network, it round-robins through the available input ports for each layer, dividing the traffic amongst the layers evenly.

Ring-Splitting: Given a 400mm^2 64-tile PNoC system, each tile is physically able to contain 80 double-back waveguides. However, the ring-splitting factor is limited to 4 (32 wavelengths per waveguide) in this design to avoid the unnecessary optical splitting loss due to the current technology. This implies a trade off of waveguide area for lower power. The splitting loss has been included in the power model in Section 6.5.

Scaling to Larger Networks: We note, it is likely that increasing cores connected in a given subnet will yield increased contention. A power-efficient means to cover the increase in bandwidth demand due to more nodes would be to increase the number of layers. We find the degree of subnet partitioning is more dependent upon the physical chip dimensions than the number of nodes connected, as the size of the chip determines the latency and frequency of arbitration phases. For this reason our base implementation assumes a large, 400mm^2 die. Increasing nodes while retain-

ing the same physical dimensions will cause a sub-linear increase in arbitration flag size with nodes-per-subnet (the Source ID would increase linearly, the Destination ID would increase as $log(n)$), and hence more overhead than in a smaller sub-net design.

6.2 Experimental Methodology

To evaluate this implementation's performance, we use a cycle-accurate, micro-architectural network simulator, ocin_tsim [50]. The network was simulated under both synthetic and realistic workloads. LumiNOC designs with 1, 2, and 4 **Network Layers** are simulated to show results for different bandwidth design points.

Photonic Networks: The baseline, 64-node LumiNOC system, as described in Section 6, was simulated for all evaluation results. Synthetic benchmark results for the Clos LTBw network are presented for comparison against the LumiNOC design. We chose the Clos LTBw design as the most competitive in terms of efficiency and bandwidth as discussed in Section 6. Clos LTBw data points were extracted from the paper by Joshi et al [15].

Baseline Electrical Network: In the results that follow, our design is compared to a electrical 2-D mesh network. Traversing the dimension order network consumes three cycles per hop; one cycle for link delay and two within each router. The routers have two virtual channels per port, each ten flits deep, and implement wormhole flow control.

Workloads: Both synthetic and realistic workloads were simulated. The traditional synthetic traffic patterns, *uniform random* and *bit-complement* represent nominal and worst-case traffic for this design. These patterns were augmented with the *P8D* pattern, proposed by Joshi et al. [15], designed as a best-case for staged or hierarchical networks where traffic is localized to individual regions. In P8D, nodes are assigned to one of 8 groups, made up of topologically adjacent nodes and nodes only send random traffic within the group. In these synthetic workloads, all packets contain data payloads of 512-bits, representing four flits of data in the baseline electrical NoC.

Realistic workload traces were captured for a 64-core CMP running PARSEC benchmarks with the sim-large input set [51]. The Netrace trace dependency tracking infrastructure was used to ensure realistic packet interdependencies are expressed as in a true, full-system CMP system [52]. The traces were captured from a CMP composed of 64 in-order cores with 32-KB, private L1I and L1D caches and a shared 16MB LLC. Coherence among the L1 caches was maintained via a MESI protocol. A 150 million cycle segment of the PARSEC benchmark "region of interest" was simulated. Packet sizes for realistic workloads vary bimodally between 64 and 576 bits for miss request/coherence traffic and cache line transfers.

6.3 Synthetic Workload Results

In Figure 11, the LumiNOC design is compared against the electrical and Clos networks under *uniform random*, *bit complement*, and *P8D*. The figure shows the low-load latencies of the LumiNOC design are much lower than the competing designs. This is due primarily to the lower diameter of the LumiNOC topology; destinations within one subnet are one "hop" away while those in a second subnet are two. The 1-layer network saturates at 4Tbps realistic throughput as determined by analyzing the offered vs. accepted rate.

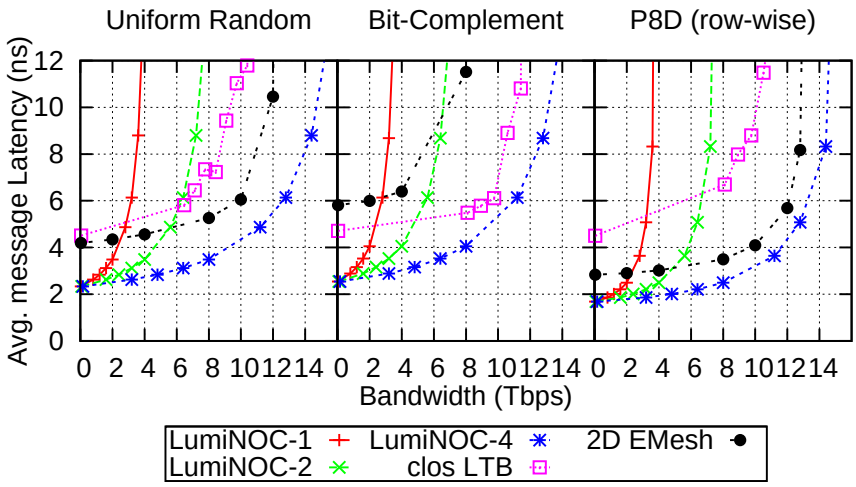


Fig. 11: Synthetic workloads showing LumiNOC vs. Clos LTBw and electrical network. LumiNOC-1 refers to the 1-layer LumiNOC design, LumiNOC-2 the 2-layer, and LumiNOC-4 the 4-layer.

The different synthetic traffic patterns bring out interesting relationships. On the *P8D* pattern, which is engineered to have lower hop counts, all designs have universally lower latency than on other patterns. However, while both the electrical and LumiNOC network have around 25% lower low-load latency than uniform random, Clos only benefits by a few percent from this optimal traffic pattern. At the other extreme, the electrical network experiences a 50% increase in no-load latency under the bit-complement pattern compared to uniform random while both Clos and the LumiNOC network are only marginally affected. This is due to the LumiNOC having a worst-case hop count of two and not all routes go through the central nodes as in the electrical network. Instead, the intermediate nodes are well distributed through the network under this traffic pattern. However, as the best-case hop count is also two with this pattern, the LumiNOC network experiences more contention and the saturation bandwidth is decreased as a result.

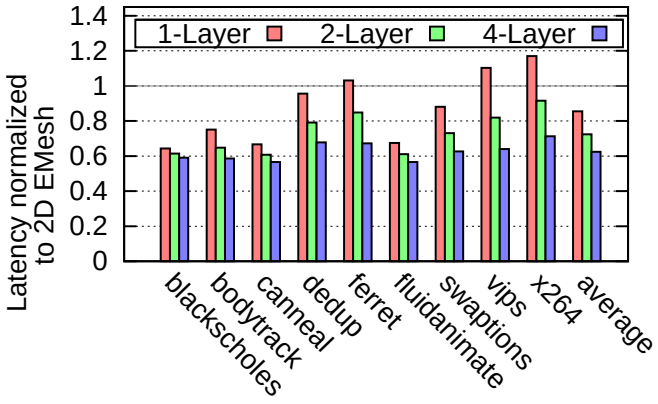


Fig. 12: Message Latency in PARSEC benchmarks for LumiNOC compared to electrical network.

Loss Component	Value	Loss Component	Value
Coupler	1 dB	Waveguide	1 dB/cm
Splitter	0.2 dB	Waveguide Crossing	0.05 dB
Non-linearity	1 dB	Ring Through	0.001 dB
Modulator Insertion	0.001 dB	Filter Drop	1.5 dB
Photodetector	0.1 dB		

Table 1: Components of optical loss.

6.4 Realistic Workload Results

Figure 12 shows the performance of the LumiNOC network in 1-, 2- and 4-layers, normalized against the performance of the baseline electrical NoC. Even with one layer, the average message latency is about 10% lower than the electrical network. With additional network layers, LumiNOC has approximately 40% lower average latency. These results are explained by examining the bandwidth-latency curves in Figure 11. The average offered rates for the PARSEC benchmarks are of the order of 0.5Tbps, so these applications benefit from LumiNOC's low latency while being well under even the 1-layer LumiNOC throughput.

6.5 PNoC Power Model

In this section, we describe our power model and compare the baseline LumiNOC design against prior work PNoC architectures. In order for a fair comparison versus other reported PNoC architectures, we refer to the photonic loss of various photonic devices reported by Joshi et al. [15] and Pan et al. [18], shown in Table 1. Equation 1

Literature		N_{core}	N_{node}	N_{rt}	N_{wg}	N_{wv}	N_{ring}
EMesh [1]		128	64	64	NA	NA	NA
Corona [21]		256	64	64	388	24832	1056K
FlexiShare [18]		64	32	32	NA	2464	550K
Clos [15]		64	8	24	56	3584	14K
LumiNOC	1-L	64	64	64	32	1024	16K
	2-L	64	64	64	64	2048	32K
	4-L	64	64	64	128	4096	65K

Table 2: Configuration comparison of various photonic NoC architectures - N_{core} : number of cores in the CMP, N_{node} : number of nodes in the NoC, N_{rt} : total number of routers, N_{wg} : total number of waveguides, N_{wv} : total number of wavelengths, N_{ring} : total number of rings.

shows the major components of our total power model.

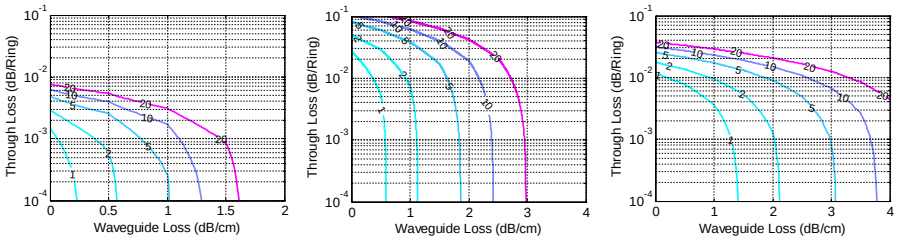
$$TP = ELP + TTP + ERP + EO/OE \tag{1}$$

TP = Total Power, ELP = Electrical Laser Power, TTP = Thermal Tuning Power, ERP = Electrical Router Power and EO/OE = Electrical to Optical/Optical to Electrical conversion power. Each components is described below.

ELP: Electrical laser power is converted from the calculated optical power. Assuming a 10μW receiver sensitivity, the minimum static optical power required at each wavelength to activate the last photodetector at the end of a waveguide in the PNoC system is estimated based on Equation 2. This optical power is then converted to electrical laser power using 30% efficiency.

$$P_{optical} = N_{wg} \cdot N_{wv} \cdot P_{th} \cdot K \cdot 10^{(\frac{1}{10} \cdot l_{channel} \cdot P_{WG\_loss})} \cdot 10^{(\frac{1}{10} \cdot N_{ring} \cdot P_{t\_loss})} \tag{2}$$

In Equation 2, N_{wg} is the number of waveguides in the PNoC system, N_{wv} is the number of wavelength per waveguide, P_{th} is receiver sensitivity power, $l_{channel}$ is waveguide length, P_{wg_loss} is optical signal propagation loss in waveguide (dB / cm), N_{ring} is the number of rings attached on each waveguide, P_{t_loss} is modulator insertion and filter ring through loss (dB / ring) (assume they are equal), K accounts for the other loss components in the optical path including P_c, coupling loss between the laser source and optical waveguide, P_b, waveguide bending loss, and $P_{splitter}$, optical splitter loss. Figure 13 shows electrical laser power contour plot, derived from Equation 2 and the configurations of Table 2, showing the photonic device power requirements at a given electrical laser power, for a SWMR photonic crossbar (Corona) [21], Clos [15] and LumiNOC with equivalent throughput (20Tbps), network radix and chip area. In Figure 13, the x and y-axis represent two major optical loss components, waveguide propagation loss and ring through loss, respectively. A larger x- and y-intercept implies relaxed requirements for the photonic devices. As shown, given a relatively low 1W laser power budget, the two-layer LumiNOC can operate with a maximum 0.012dB ring through loss and waveguide loss of 1.5dB/cm.



(a) Crossbar (b) Clos (c) LumiNOC

Fig. 13: Contour plots of the Electrical Laser Power (ELP) in Watts for networks with the same aggregate throughput. Each line represents a constant power level (Watts) at a given ring through loss and waveguide loss combination (assuming 30% efficient electrical to optical power conversion).

We note that optical non-linear loss also effects the optical interconnect power. At telecom wavelengths, two-photon absorption (TPA) in the silicon leads to a propagation loss that increases linearly with the power sent down the waveguide. TPA is a nonlinear optical process and is several orders of magnitude weaker than linear absorption. This nonlinear loss, however, also has significant impact on the silicon-photonic link power budget if a high level of optical power (e.g. $>$1W) is injected into silicon waveguide. Figure 14 shows the computed nonlinear loss of a 1cm waveguide versus the optical power in the waveguide. It shows a nonlinear loss of $\sim$0.35 dB for up to $\sim$100 mW waveguide optical power. In LumiNoC, the non-linear effect has been included in the optical power calculation.

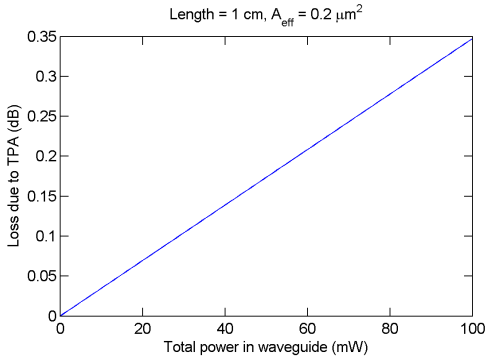


Fig. 14: Nonlinear optical loss in the silicon waveguide vs optical power in waveguide; waveguide length equals 1cm with effective area of $0.2um^2$. Figure produced by Jason Pelc of HP labs with permission.

Literature		ELP (W)	TTP (W)	ERP (W)	EO/OE (W)	ITP (Tbps)	RTP (Tpbs)	TP (W)	**RTP/W (Tbps/W)**
EMesh [1]		NA	NA	NA	NA	10	3.0	26.7	**0.1**
Corona [21]		26.0	21.00	0.52	4.92	160	73.6	52.4	**1.4**
FlexiShare [18]		5.80	11.00	0.13	0.60	20	9.0	17.5	**0.5**
Clos [15]		3.30	0.14	0.10	0.54	18	10.0	4.1	**2.4**
LumiNOC	1-Layer	0.35	0.33	0.13	0.30	10	4.0	1.1	**3.6**
	2-Layers	0.73	0.65	0.26	0.61	20	8.0	2.3	**3.4**
	4-Layers	1.54	1.31	0.52	1.22	40	16.0	4.6	**3.4**

Table 3: Power efficiency comparison of different photonic NoC architectures - ELP : Electrical Laser Power, TTP : Thermal Tuning Power, ERP : Electrical Router Power, EO/OE : Electrical to optical/Optical to electrical conversion power, ITP : Ideal Throughput, RTP : Realistic Throughput, TP : Total Power.

TTP: Thermal tuning is required to maintain microring resonant at the work wavelength. In the calculation, a ring thermal tuning power of 20μW is assumed for a 20K temperature tuning range [15, 18]. In a photonic NoC, total thermal tuning power (TTP) is proportional to ring count.
ERP: The baseline electrical router power is estimated by the power model reported by Kim et al. [53]. We synthesized the router using TSMC 45nm library. Power is measured via Synopsis Power Compiler, using simulated traffic from a PARSEC [51] workload to estimate its dynamic component. Results are analytically scaled to 22nm (dynamic power scaled according to the CMOS dynamic power equation and static power linearly with voltage).
EO/OE: The power for conversion between the electrical and optical domains (EO/OE) is based on the model reported by Joshi et al. [15], which assumes a total transceiver energy of 40 fJ/bit data-traffic dependent energy and 10 fJ/bit static energy. Since previous photonic NoCs consider different traffic loads, it is unfair to compare the EO/OE power by directly using their reported figures. Therefore, we compare the worst-case power consumption when each node was arbitrated to get a full access on each individual channel. For example, Corona is a MWSR 64×64 crossbar architecture. At the worst-case, 64 nodes are simultaneously writing on 64 different channels. This is combined with a per-bit activity factor of 0.5 to represent random data in the channel.

While this approach may not be 100% equitable for all designs, we note that EO/OE power does not dominate in any of the designs (see Table 3). Even if EO/OE power is removed entirely from the analysis, the results would not change significantly. Further, LumiNOC experiences more EO/OE dynamic power than the other designs due hops through the middle routers.

6.6 Power Comparison

Table 2 lists the photonic resource configurations for various photonic NoC architectures, including one-layer, two-layer and four-layer configurations of the LumiNOC.

While the crossbar architecture of Corona has a high ideal throughput, the excessive number of rings and waveguides results in degraded power efficiency. In order to support equal 20Tbps aggregate throughput, LumiNOC requires less than $\frac{1}{10}$ the number of rings of FlexiShare and almost the same number of wavelengths. Relative to the Clos architecture, LumiNOC requires around $\frac{4}{7}$ wavelengths, though approximately double number of rings.

The power and efficiency of the network designs is compared in Table 3. Where available/applicable, power and throughput numbers for competing PNoC designs are taken from the original papers, otherwise they are calculated as described in Section 6.5. **ITP** is the ideal throughput of the design, while **RTP** is the maximum throughput of the design under a uniform random workload as shown in Figure 11. A 6×4 2GHz electrical 2D-mesh [1] was scaled to 8×8 nodes operating at 5GHz, in a 22nm CMOS process (dynamic power scaled according to the CMOS dynamic power equation and static power linearly with voltage), to compare against the photonic networks.

The table shows that LumiNOC has the highest power efficiency of all designs compared in RTP/Watt, increasing efficiency by ∼40% versus the nearest competitor, Clos [15]. By reducing wavelength multiplexing density, utilizing shorter waveguides, and leveraging the data channels for arbitration, LumiNOC consumes the least ELP among all the compared architectures. A 4-layer LumiNOC consumes ∼1/4th the ELP of a competitive Clos architecture, of nearly the same throughput. Corona [21] contains 256 cores with 4 cores sharing an electrical router, leading to a 64-node photonic crossbar architecture; however, in order to achieve throughput of 160Gbps, each channel in Corona consists of 256 wavelengths, 4X the wavelengths in a 1-layer LumiNOC. In order to support the highest ideal throughput, Corona consumes the highest electrical router power in the compared photonic NoCs.

Although FlexiShare attempts to save laser power with its double-round waveguide, which reduces the overall non-resonance ring through-loss (and it is substantially more efficient than Corona), its RTP/W remains somewhat low for several reasons. First, similar to other PNoC architectures, FlexiShare employs a global, long waveguide bus instead of multiple short waveguides for the optical interconnects. The global long waveguides cause relatively large optical loss and overburden the laser. Second, FlexiShare is particularly impacted by the high number of ring resonators ($N_{ring} = 550K$ - Table 2), each of these rings need to be heated to maintain its proper frequency response and the power consumption of this heating dominates its RTP/W. Third, the dedicated physical arbitration channel in FlexiShare costs extra optical power. Finally, similar to an SWMR crossbar network (e.g. Firefly [19]), FlexiShare broadcasts to all the other receivers for receiver-side arbitration. Although the authors state that, by only broadcasting the head flit, the cost of broadcast in laser power is avoided, we would argue this would be impractical in practice. Since the turn-around time for changing off-die laser power is so high, a constant laser power is needed to support the worst-case power consumption.

7 Conclusions

Photonic NoCs are a promising replacement for electrical NoCs in future many-core processors. In this work, we analyze prior photonic NoCs, with an eye towards efficient system power utilization and low-latency. The analysis of prior photonic NoCs reveals that power inefficiencies are mainly caused by channel over-provisioning, unnecessary optical loss due to topology and photonic device layout and power overhead from the separated arbitration channels and networks. LumiNOC addresses these issues by adopting a shared-channel, photonic on-chip network with a novel, in-band arbitration mechanism to efficiently utilize power, achieving a high performance and scalable interconnect with extremely low latency. Simulations show under synthetic traffic, LumiNOC enjoys 50% lower latency at low loads and $\sim$40% higher throughput per Watt on synthetic traffic, versus other reported photonic NoCs. LumiNOC also reduces latencies $\sim$40% versus an electrical 2D mesh NoCs on the PARSEC shared-memory, multithreaded benchmark suite.

References

[1] J. Howard, S. Dighe, S. R. Vangal, G. Ruhl, N. Borkar, S. Jain, V. Erraguntla, M. Konow, M. Riepen, M. Gries, G. Droege, T. Lund-Larsen, S. Steibl, S. Borkar, V. K. De, and R. V. D. Wijngaart, "A 48-Core IA-32 Processor in 45 nm CMOS Using On-Die Message-Passing and DVFS for Performance and Power Scaling," *IEEE Journal of Solid-State Circuits*, vol. 46, pp. 173–183, Oct 2011.

[2] M. Anders, H. Kaul, S. Hsu, A. Agarwal, S. Mathew, F. Sheikh, R. Krishnamurthy, and S. Borkar, "A 4.1tb/s bisection-bandwidth 560gb/s/w streaming circuit-switched mesh network-on-chip in 45nm cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, pp. 110–111, Feb 2010.

[3] J. Kim, D. Park, T. Theocharides, N. Vijaykrishnan, and C. R. Das, "A Low Latency Router Supporting Adaptivity for On-Chip Interconnects," in *2005 Design Automation Conference*, pp. 559–564, June 2005.

[4] M. Lipson, "Compact electro-optic modulators on a silicon chip," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 12, no. 6, pp. 1520–1526, 2006.

[5] A. Liu, L. Liao, D. Rubin, H. Nguyen, B. Ciftcioglu, Y. Chetrit, N. Izhaky, and M. Paniccia, "High-speed optical modulation based on carrier depletion in a silicon waveguide," *Optics Express*, vol. 15, no. 2, pp. 660–668, 2007.

[6] M. Reshotko, B. Block, B. Jin, and P. Chang, "Waveguide Coupled Ge-on-Oxide Photodetectors for Integrated Optical Links," in *The 2008 5th IEEE International Conference on Group IV Photonics*, pp. 182–184, 2008.

[7] C. Holzwarth, J. Orcutt, H. Li, M. Popovic, V. Stojanovic, J. Hoyt, R. Ram, and H. Smith, "Localized Substrate Removal Technique Enabling Strong-

Confinement Microphotonics in Bulk Si CMOS Processes," in *Conference on Lasers and Electro-Optics*, pp. 1–2, 2008.

[8] L. C. Kimerling, D. Ahn, A. Apsel, M. Beals, D. Carothers, Y.-K. Chen, T. Conway, D. M. Gill, M. Grove, C.-Y. Hong, M. Lipson, J. Michel, D. Pan, S. S. Patel, A. T. Pomerene, M. Rasras, D. K. Sparacin, K.-Y. Tu, A. E. White, and C. W. Wong, "Electronic-photonic integrated circuits on the CMOS platform," in *Silicon Photonics*, pp. 6–15, 2006.

[9] A. Narasimha, B. Analui, Y. Liang, T. Sleboda, and C. Gunn, "A Fully Integrated 4 10-Gb/s DWDM Optoelectronic Transceiver Implemented in a Standard 0.13 m CMOS SOI Technology," in *The IEEE International Solid-State Circuits Conference*, pp. 42–586, 2007.

[10] I. Young, E. Mohammed, J. Liao, A. Kern, S. Palermo, B. Block, M. Reshotko, and P. Chang, "Optical I/O Technology for Tera-scale Computing," in *The IEEE International Solid-State Circuits Conference*, pp. 468–469, 2009.

[11] G. Hendry, S. Kamil, A. Biberman, J. Chan, B. Lee, M. Mohiyuddin, A. Jain, K. Bergman, L. Carloni, J. Kubiatowicz, L. Oliker, and J. Shalf, "Analysis of Photonic Networks for a Chip Multiprocessor using Scientific Applications," in *The 3rd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, pp. 104–113, 2009.

[12] A. Shacham, K. Bergman, and L. P. Carloni, "On The Design of a Photonic Network-On-Chip," in *The First International Symposium on Networks-on-Chip (NOCS)*, pp. 53–64, 2007.

[13] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic NoC for DMA Communications in Chip Multiprocessors," in *The 15th Annual IEEE Symposium on High-Performance Interconnects*, pp. 29–38, 2007.

[14] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic Networks-On-Chip for Future Generations of Chip Multiprocessors," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1246–1260, 2008.

[15] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic, "Silicon-Photonic Clos Networks for Global On-Chip Communication," in *The 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip (NOCS)*, pp. 124–133, 2009.

[16] N. Kirman, M. Kirman, R. Dokania, J. Martinez, A. Apsel, M. Watkins, and D. Albonesi, "Leveraging Optical Technology in Future Bus-Based Chip Multiprocessors," in *The 39th Annual IEEE/ACM International Symposium on Microarchitecture (Micro)*, pp. 492–503, 2006.

[17] A. Krishnamoorthy, R. Ho, X. Zheng, H. Schwetman, J. Lexau, P. Koka, G. Li, I. Shubin, and J. Cunningham, "Computer Systems Based on Silicon Photonic Interconnects," *Proceedings of the IEEE*, vol. 97, no. 7, pp. 1337–1361, 2009.

[18] Y. Pan, J. Kim, and G. Memik, "FlexiShare: Channel Sharing for an Energy-Efficient Nanophotonic Crossbar," in *The 16th IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pp. 1–12, 2010.

[19] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary, "Firefly: Illuminating future network-on-chip with nanophotonics," in *36th International Symposium on Computer Architecture (ISCA)*, 2009.

[20] D. Vantrease, N. Binkert, R. Schreiber, and M. H. Lipasti, "Light Speed Arbitration and Flow Control for Nanophotonic Interconnects," in *42nd Annual IEEE/ACM International Symposium on microarchitecture*, pp. 304–315, 2009.

[21] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn, "Corona: System Implications of Emerging Nanophotonic Technology," in *35th International Symposium on Computer Architecture (ISCA)*, pp. 153–164, 2008.

[22] P. Koka, M. O. McCracken, H. Schwetman, X. Zheng, R. Ho, and A. V. Krishnamoorthy, "Silicon-Photonic Network Architectures for Scalable, Power-Efficient Multi-Chip Systems," in *37th International Symposium on Computer Architecture (ISCA)*, pp. 117–128, 2010.

[23] Y. H. Kao and H. J. Chao, "BLOCON: A Bufferless Photonic Clos Network-on-Chip Architecture," in *5th ACM/IEEE International Symposium on Networks-on-Chip (NoCS)*, pp. 81–88, May 2011.

[24] C. Li, M. Browning, P. V. Gratz, and S. Palermo, "Luminoc: A power-efficient, high-performance, photonic network-on-chip for future parallel architectures," in *Proceedings of the 21st International Conference on Parallel Architectures and Compilation Techniques*, PACT '12, (New York, NY, USA), pp. 421–422, ACM, 2012.

[25] I. Young, E. Mohammed, J. Liao, A. Kern, S. Palermo, B. Block, M. Reshotko, and P. Chang, "Optical I/O technology for tera-scale computing," *IEEE Journal of Solid-State Circuits*, vol. 45, pp. 235–248, Jan 2010.

[26] B. G. Lee, A. V. Rylyakov, W. M. J. Green, S. Assefa, C. W. Baks, R. Rimolo-Donadio, D. M. Kuchta, M. H. Khater, T. Barwicz, C. Reinholm, E. Kiewra, S. M. Shank, C. L. Schow, and Y. A. Vlasov, "Four- and eight-port photonic switches monolithically integrated with digital CMOS logic and driver circuits," *IEEE-OSA Optical Fiber Communications Conference*, pp. 1–3, Mar 2013.

[27] J. E. Roth, S. Palermo, N. C. Helman, D. P. Bour, D. A. B. Miller, and M. Horowitz, "An optical interconnect transceiver at 1550nm using low-voltage electroabsorption modulators directly integrated to CMOS," *IEEE-OSA Journal of Lightwave Technology*, vol. 25, pp. 3739–3747, Dec 2007.

[28] A. Liu, L. Liao, D. Rubin, J. Basak, H. Nguyen, Y. Chetrit, R. Cohen, N. Izhaky, and M. Paniccia, "High-speed silicon modulator for future vlsi interconnect," in *Integrated Photonics and Nanophotonics Research and Applications / Slow and Fast Light*, p. IMD3, Optical Society of America, 2007.

[29] G. L. Wojcik, D. Yin, A. R. Kovsh, A. E. Gubenko, I. L. Krestnikov, S. S. Mikhrin, D. A. Livshits, D. A. Fattal, M. Fiorentino, and R. G. Beausoleil, "A single comb laser source for short reach WDM interconnects," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7230 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Feb. 2009.

[30] R. A. Soref and B. Bennett, "Electrooptical effects in silicon," *Quantum Electronics, IEEE Journal of*, vol. 23, pp. 123–129, Jan 1987.

[31] C. Li, R. Bai, A. Shafik, E. Tabasy, G. Tang, C. Ma, C.-H. Chen, Z. Peng, M. Fiorentino, P. Chiang, and S. Palermo, "A ring-resonator-based silicon photonics transceiver with bias-based wavelength stabilization and adaptive-power-sensitivity receiver," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 124–125, Feb 2013.

[32] G. Li, X. Zheng, J. Yao, H. Thacker, I. Shubin, Y. Luo, K. Raj, J. E. Cunningham, and A. V. Krishnamoorthy, "High-efficiency 25Gb/s CMOS ring modulator with integrated thermal tuning," *8th IEEE Intentional Conference on Group IV Photonics (GFP)*, vol. 4, pp. 8–10, 2011.

[33] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, "12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators," *Opt. Express*, vol. 15, pp. 430–436, Jan 2007.

[34] C.-H. Chen, C. Li, A. Shafik, M. Fiorentino, P. Chiang, S. Palermo, and R. Beausoleil, "A wdm silicon photonic transmitter based on carrier- injection microring modulators," in *Optical Interconnects Conference, 2014 IEEE*, May 2014.

[35] F. Liu, D. Patil, J. Lexau, P. Amberg, M. Dayringer, J. Gainsley, H. Moghadam, X. Zheng, J. Cunningham, A. Krishnamoorthy, E. Alon, and R. Ho, "10-gbps, 5.3-mw optical transmitter and receiver circuits in 40-nm cmos," *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 9, pp. 2049–2067, 2012.

[36] C. Sun, E. Timurdogan, M. Watts, and V. Stojanovic, "Integrated microring tuning in deep-trench bulk cmos," in *Optical Interconnects Conference, 2013 IEEE*, pp. 54–55, 2013.

[37] J. S. Orcutt, B. Moss, C. Sun, J. Leu, M. Georgas, J. Shainline, E. Zgraggen, H. Li, J. Sun, M. Weaver, S. Urošević, M. Popović, R. J. Ram, and V. Stojanović, "Open foundry platform for high-performance electronic-photonic integration," *Opt. Express*, vol. 20, pp. 12222–12232, May 2012.

[38] P. Dong, W. Qian, H. Liang, R. Shafiiha, D. Feng, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and M. Asghari, "Thermally tunable silicon racetrack resonators with ultralow tuning power," *Opt. Express*, vol. 18, pp. 20298–20304, Sep 2010.

[39] J. S. Orcutt, A. Khilo, C. W. Holzwarth, M. A. Popović, H. Li, J. Sun, T. Bonifield, R. Hollingsworth, F. X. Kärtner, H. I. Smith, V. Stojanović, and R. J. Ram, "Nanophotonic integration in state-of-the-art cmos foundries," *Opt. Express*, vol. 19, pp. 2335–2346, Jan 2011.

[40] A. Biberman, K. Preston, G. Hendry, N. Sherwood-droz, J. Chan, J. S. Levy, M. Lipson, and K. Bergman, "Photonic Network-on-Chip Architectures Using Multilayer Deposited Silicon Materials for High-Performance Chip Multiprocessors," *ACM Journal on Emerging Technologies in Computing Systems*, vol. 7, no. 2, pp. 1305–1315, 2011.

[41] G. Hendry, E. Robinson, V. Gleyzer, J. Chan, L. P. Carloni, N. Bliss, and K. Bergman, "Time-Division-Multiplexed Arbitration in Silicon Nanophotonic Networks-On-Chip for High-Performance Chip Multiprocessors," *Journal of Parallel and Distributed Computing*, vol. 71, pp. 641–650, May 2011.

[42] C. Chen and A. Joshi, "Runtime management of laser power in silicon-photonic multibus noc architecture," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 19, no. 2, 2013.

[43] L. Zhou and A. Kodi, "Probe: Prediction-based optical bandwidth scaling for energy-efficient nocs," in *Networks on Chip (NoCS), 2013 Seventh IEEE/ACM International Symposium on*, pp. 1–8, 2013.

[44] A. Kodi and R. Morris, "Design of a scalable nanophotonic interconnect for future multicores," in *The 5th ACM/IEEE Symposium on Architectures for Networking and Communications Systems*, pp. 113–122, ACM, 2009.

[45] R. W. Morris and A. K. Kodi, "Power-Efficient and High-Performance Multi-Level Hybrid Nanophotonic Interconnect for Multicores," in *4th ACM/IEEE International Symposium on Networks-on-Chip (NoCS)*, pp. 207–214, May 2010.

[46] S. Bahirat and S. Pasricha, "Uc-photon: A novel hybrid photonic network-on-chip for multiple use-case applications," in *Quality Electronic Design (ISQED), 2010 11th International Symposium on*, pp. 721–729, IEEE, 2010.

[47] J. Xue, A. Garg, B. Ciftcioglu, J. Hu, S. Wang, I. Savidis, M. Jain, R. Berman, P. Liu, M. Huang, H. Wu, E. Friedman, G. Wicks, and D. Moore, "An intra-chip free-space optical interconnect," in *Proceedings of the 37th annual international symposium on Computer architecture*, ISCA '10, (New York, NY, USA), pp. 94–105, ACM, 2010.

[48] P. Gratz and S. W. Keckler, "Realistic Workload Characterization and Analysis for Networks-on-Chip Design," in *The 4th Workshop on Chip Multiprocessor Memory Systems and Interconnects (CMP-MSI)*, 2010.

[49] M. R. T. Tan, P. Rosenberg, S. Mathai, J. Straznicky, L. Kiyama, J. S. Yeo, M. Mclaren, W. Mack, P. Mendoza, and H. P. Kuo, "Photonic Interconnects for Computer Applications," in *Communications and Photonics Conference and Exhibition (ACP), 2009 Asia*, pp. 1–2, 2009.

[50] S. Prabhu, B. Grot, P. Gratz, and J. Hu, "Ocin tsim-DVFS Aware Simulator for NoCs," *Proc. SAW*, vol. 1, 2010.

[51] C. Bienia, S. Kumar, J. P. Singh, and K. Li, "The PARSEC Benchmark Suite: Characterization and Architectural Implications," in *The 17th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, October 2008.

[52] J. Hestness and S. Keckler, "Netrace: Dependency-Tracking Traces for Efficient Network-on-Chip Experimentation," tech. rep., Technical Report TR-10-11, The University of Texas at Austin, Department of Computer Science, http://www.cs.utexas.edu/~netrace, 2010.

[53] H. Kim, P. Ghoshal, B. Grot, P. V. Gratz, and D. A. Jimenez, "Reducing network-on-chip energy consumption through spatial locality speculation," in *5th ACM/IEEE International Symposium on Networks-on-Chip (NoCS)*, pp. 233–240, 2011.