A Low-Latency, High-Throughput On-Chip Optical Router Architecture for Future Chip Multiprocessors

MARK J. CIANCHETTI and DAVID H. ALBONESI, Cornell University

Tens and eventually hundreds of processing cores are projected to be integrated onto future microprocessors, making the global interconnect a key component to achieving scalable chip performance within a given power envelope. While CMOS-compatible nanophotonics has emerged as a leading candidate for replacing global wires beyond the 16nm timeframe, on-chip optical interconnect architectures are typically limited in scalability or are dependent on comparatively slow electrical control networks.

In this article, we present a hybrid electrical/optical router for future large scale, cache coherent multicore microprocessors. The heart of the router is a low-latency optical crossbar that uses predecoded source routing and switch state preconfiguration to transmit cache-line-sized packets several hops in a single clock cycle under contentionless conditions. Overall, our optical router achieves 2X better network performance than a state-of-the-art electrical baseline in a mesh topology while consuming 30% less network power.

Categories and Subject Descriptors: C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Packet-switching networks*

General Terms: Design, Experimentation, Performance

Additional Key Words and Phrases: Interconnect architecture, multiprocessor network, optical interconnect, optical switch

ACM Reference Format:
Cianchetti, M. J. and Albonesi, D. H. 2011. A low-latency, high-throughput on-chip optical router architecture for future chip multiprocessors. ACM J. Emerg. Technol. Comput. Syst. 7, 2, Article 9 (June 2011), 20 pages.
DOI = 10.1145/1970406.1970411 http://doi.acm.org/10.1145/1970406.1970411

1. INTRODUCTION

As the microprocessor industry moves to integrating tens of cores on a single die, the global interconnect becomes a critical performance bottleneck. The ITRS Roadmap [ITRS 2008] projects that metal interconnects will become inadequate to meet the speed and power dissipation requirements of highly scaled ICs beyond 22nm and lists CMOS-compatible optical interconnects as a possible solution.

The potential advantages of optical interconnects include high-speed signal propagation, high bandwidth density through time and wavelength division multiplexing (TDM and WDM), and low crosstalk between signal paths [Miller 2000]. In recent years, significant advances in CMOS-compatible optical components [Almeida et al. 2004; Paniccia et al. 2006; Park et al. 2007; Preston et al. 2008; Tatum 2001] have brought the technology closer to commercial viability. As a result, several teams have proposed detailed architectural designs for multicore chips with integrated optical technology [Joshi et al. 2009; Kirman et al. 2006; Kirman and Martinez 2010; Shacham et al. 2007; Vantrease et al. 2008], and a number of microprocessor manufacturers are

Author's address: email: Mjc96@cornell.edu.

DOI 10.1145/1970406.1970411 http://doi.acm.org/10.1145/1970406.1970411

investigating silicon photonic devices and architectures for on-chip communication in these future systems.

However, optical technology has several drawbacks. Optical logic gates [Xu and Lipson 2007] and storage (e.g., buffers [Xi et al. 2006]) are far from mature; thus, control must be implemented in the electrical domain, and buffering likewise must be performed electrically. As a result, many proposed optical interconnect architectures are bus-based. For example, the Cornell hybrid electrical/optical interconnect architecture [Kirman et al. 2006] comprises an optical ring that assigns unique wavelengths per node in order to implement a multibus. Every bus cycle, the contents of the buses are optically received, converted to electrical signals, and then handled by logic in the electrical domain (decoded, etc.). The HP Corona crossbar architecture [Vantrease et al. 2008] is in fact numerous multiple writer, single reader buses routed in a snake pattern among the nodes.

In these two approaches, the physical topology is chosen to avoid waveguide crossings, and a bus approach prevents control functionality from limiting data transmission speed. The Columbia optical network [Shacham et al. 2007] is one of the few that proposes on-chip optical switches. The network consists of a 2D grid of optical waveguides with optical resonators at intersecting points to perform turns. An electrical subnetwork sets up the switches in advance of data transmission and tears down the network thereafter. Once the path is set up, communication proceeds between source and destination. The small number of waveguides in each channel limits the number of crossings. Moreover, the setup is done in advance so that the control circuitry does not limit transmission speed. However, the network must transmit a large amount of data to amortize the relatively high latency of the electrical setup/teardown network, making the network unsuitable for a typical cache coherent shared memory system where the unit of transfer is a cache line.

In this article, we present a hybrid optical/electrical router to facilitate high-speed cache coherent transfers in future multicore microprocessors.[1] We enable a packet to traverse multiple hops in a clock cycle through the use of low latency optical devices, predecoded source routing and localized router control, while using less power than a state-of-the-art electrical interconnect. The simplicity of the router's architecture suggests its extensibility to high and low radix switching, enabling its use in a variety of on-chip network topologies.

We expand on our original work in Phastlane [Cianchetti et al. 2009] through the complete redesign of the optical router. We present a new switch architecture which localizes all router control logic within each input port, removing any delays associated with propagation of electrical control signals. We implement optical switch arbitration, guaranteeing fairness to all packets through rotating priority arbitration. Phastlane suffered from dropped packets when a downstream router's buffering was full, wasting precious resources and energy. We remedy this problem through a unique implementation of On/Off flow control.

To achieve even lower router delay, we introduce a mechanism for preconfiguring switch state so that packets can travel even faster from an input port to an output port. This allows a packet to traverse up to 8 hops, assuming dimension-order routing, in a single clock cycle. Through detailed architectural simulations we examine the performance of our optical router in a mesh network and demonstrate an overall 2X network speedup with 30% less power consumption than a state-of-the art electrical network.

[1]This work builds on Phastlane, a hybrid optical/electrical mesh network [Cianchetti et al. 2009].

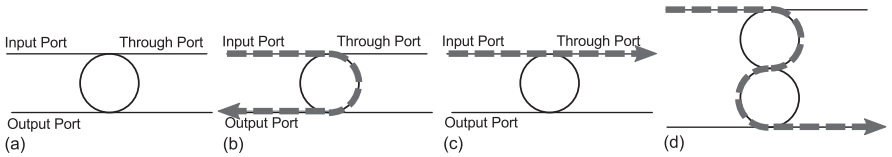


Fig. 1. (a) Microring optical resonator used as a switching element. (b) When light corresponding to a resonant frequency of the ring travels through the Input Port, it couples into the ring and exits through the Output Port. (c) Similarly, light that does not match a resonant frequency of the ring continues to the Through Port. In (d) series-coupled rings can be used to enable high bandwidth communication through dense WDM.

2. ENABLING OPTICAL TECHNOLOGY

2.1. Microring Resonator

Ring resonators are circular waveguides that form the building blocks of an optical packet switched network through their use as optical modulators and filters. Figure 1(a) shows a ring resonator coupled between two adjacent waveguides. The material properties and geometry of the resonator determine its resonant frequencies. In Figure 1(b) when a light source enters through the Input Port, wavelengths that match a resonant frequency of the ring will couple and leave via the Output Port. Wavelengths of light which do not couple continue to the Through Port as shown in Figure 1(c). The resonant frequencies of the ring are cyclical and the distance between peaks is known as the free spectral range (FSR). The FSR spacing of ring modulators and single wavelength filters dictates the degree of wavelength-division-multiplexing (WDM), which allows multiple wavelengths to simultaneously traverse a single waveguide. Decreasing the radii of these rings will increase this spacing, enabling more dense WDM, but at the cost of increased fabrication complexity and optical power loss in the resonator [Barwicz et al. 2004].

Two ways to increase the FSR without having to fabricate ultracompact rings are through the Vernier Effect [Yanagase et al. 2002] and series-coupled resonators [Barwicz et al. 2004]. It is possible to align multiple ring resonators like in Figure 1(d) to mimic the frequency response of a single smaller ring. The output response of the full device is the product of the filter responses of each individual resonator. In this work we utilize this technique to implement ring modulators and single wavelength filters, enabling a high degree of WDM [Joshi et al. 2009].

In network applications it is desirable to switch all of the wavelengths in a WDM waveguide simultaneously using the same device, in contrast to modulators and single wavelength filters that handle a single wavelength. This is because a significant amount of data traverses between a source and destination along the same path. Increasing the radii of a ring resonator will shorten its FSR, facilitating broadband switching (known as Comb switching) thus enabling this functionality [Xia et al. 2006].

The resonant frequencies of a ring are fixed at fabrication but can be changed dynamically through the free carrier dispersion effect [Xu et al. 2007]. If carriers are injected into the ring, its effective index is changed, which causes a shift in its resonant frequencies. Interconnect networks require some level of control to set up and propagate data to a destination. A ring resonator achieves this functionality by coupling light only when a controlling electrical circuit turns it on.

2.2. Optical Waveguides

In this study we opt to use Si_3N_4 (silicon nitride) waveguides to route optical data packets between source and destination network nodes. Unlike traditional silicon-on-insulator (SOI) waveguides, Si_3N_4 compounds do not have to be grown from a seed

layer, and can be deposited directly above an active electronic die. While these devices tend to be slightly larger in size, they present a smaller propagation loss to optical signals. Silicon waveguides generally exhibit 1–3dB/cm, where nitride enables .1dB/cm propagation loss [Gondarenko et al. 2009; Shaw et al. 2005; Melchiorri et al. 2005].

Traditionally, a limiting factor in the degree of WDM was optical power loss due to nonlinear attenuation in silicon when a large amount of power travels through the waveguide. The wide bandgap of silicon nitride waveguides eliminates this drawback, allowing hundreds of wavelengths to be used for high bandwidth communication [Gondarenko et al. 2009]. Lastly, we utilize polycrystalline silicon for our resonators [Preston et al. 2008, 2007] and polycrystalline germanium detectors [Young et al. 2010], which like Si_3N_4 can both be deposited above the processor die.

2.3. Multiple Layer Waveguide Integration

Deposited materials used to construct the optical components enable multiple layer integration to eliminate waveguide crossings. Previous work [Dokania and Apsel 2009] has proposed the use of two optical waveguide layers which can be fabricated such that coupling between them is accomplished through the use of a ring resonator. While the optical losses associated with a single waveguide crossing have been demonstrated to be small [Liu et al. 2004; Bogaerts et al. 2007], achieving approximately 98% coupling efficiency, losses can accumulate when traversing many hops. In this work we assume optical waveguide layers to avoid such losses.

3. RELATED WORK

Several on-chip interconnect architectures have been proposed that leverage CMOS-compatible photonics for future multicore microprocessors. Kirman et al. [2006] propose a hierarchical interconnect for communication among 64 cores in 32nm technology. A group of four cores and a shared L2 cache communicates with four other groups through an electrical switch. The four 16-processor nodes in turn communicate using an optical ring that implements a bus protocol. Each node writes to the bus using its own unique wavelengths, which obviates the need for arbitration, and information is read by coupling a percentage of the power from each signal.

Vantrease et al. [2008] also propose optical buses for communication among 256 cores in 16nm technology. Similar to Kirman et al. [2006], multiple cores are grouped as a node and communicate through an electrical subnetwork. Internode communication occurs through a set of multiple-writer, single-reader buses (one for each node) that together form a crossbar. Optical arbitration resolves conflicts for writing a given bus. An optical token travels around a special arbitration waveguide, and a node reads and removes the token before communicating with its intended target. Chip-to-chip serial optical links communicate with main memory modules that are divided among the network nodes.

Perhaps the closest work to ours is the optical 2D network proposed by Shacham et al. [2007]. Data transfer occurs through a grid of waveguides with resonators at crosspoints for turns. Control is handled by an electrical set-up/tear-down network. To enable data transfer, a packet is sent on the electrical network which moves toward the destination and reserves the optical switches along its route. When this path is established, the source transfers data at high bandwidth using the optical network. Finally, a packet is sent in the electrical network to tear down the established path.

Kumar et al. [2007] propose Express Virtual Channels to reduce packet latency in an electrical router beyond techniques such as lookahead routing and speculation [Peh and Dally 2001]. Packets within these channels can bypass the router pipeline.

Our approach leverages elements of each of these prior proposals. Like Shacham et al., we use a grid of waveguides with turn resonators, but there are several important

distinctions between our proposals, some of which are due to differences in data payload size. We rely on only WDM to pack a narrow packet into one cycle, while they use WDM and TDM to achieve very high bandwidth transfer of a much greater amount of data. We optically send control along with the data to set up the router switches on the fly rather than use a slower electrical control network. Like Kirman et al. and Vantrease et al., we target snoopy cache-coherence multicore systems, but our networks are quite different (switch-based rather than bus-based). Finally, as with Express Virtual Channels, we seek to reduce packet latency but we do this without special dedicated express lanes. Rather, we use simple control to exploit the capability of optics to travel multiple hops in a single cycle.

Optical burst switching operates by transmitting variable sized data bursts behind a path setup signal which configures every switch ahead of time according to the packet's desired destination. Traditionally if a burst's control signal is unable to obtain a switch, it is dropped. Other work has examined deflection routing and delay lines to partially remedy this problem [Battestilli and Perros 2003; Chen et al. 2004]. Our packet switched optical router architecture transmits cache lined sized packets in parallel with router control signals. Additionally, packets are electrically buffered at routers at the end of a network clock cycle or under contention.

Kirman and Martinez [2010] propose a point-to-point optical network which routes transmissions through statically configured switches. The switch configurations are fixed at design time to route wavelengths between input and output ports. When a node submits a packet into the network, it transmits on particular wavelengths corresponding to its desired destination. These wavelengths route through the passive resonators in the network towards the destination. The network is laid out in a bus configuration to avoid waveguide crossings and increase bisection bandwidth.

4. PROPOSED NETWORK ARCHITECTURE

4.1. Router Microarchitecture

One advantage of on-chip silicon photonics is its low latency transmission over distances long enough to amortize the costs of modulation, detection, and conversion. In 16nm technology, the distance beyond which optics achieves lower delay than optimally repeatered wires is expected to be 1–2mm [Chen et al. 2005], making optical transmission profitable for even single hop network traversals. Our goal is to design an optical router that matches the performance of a state-of-the-art electrical switch under high load, but enables multiple hops to be traversed in a network cycle under reduced load. This is possible through simplicity in the router control path and switch preconfiguration, which allows an incoming packet to travel through a switch with minimal delay.

Our design targets cache coherent multicore processors in the 16nm generation with tens to hundreds of cores and a highly interleaved main memory using multiple on-chip memory controllers. Each node includes one or more processing cores, a two-level cache hierarchy, a memory controller (MC), and a network switch. The MC's are interleaved on a cache line basis with high bandwidth serial optical links like those proposed for Corona [Vantrease et al. 2008] connecting each MC to off-chip DRAM.

4.2. Switch Design

Figure 2 shows a portion of the optical components in our proposed radix five optical switch. Two of the five ports, one being the port to the local processor, are located on the west side of each router. Only two waveguides per port are shown for simplicity, a single input and a single output waveguide. A data path width of six waveguides is actually implemented to achieve high bandwidth, low latency network communication.

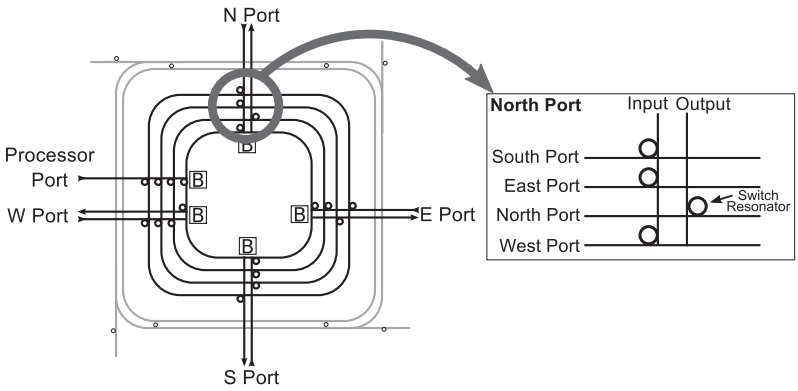


Fig. 2. Proposed optical switch architecture. The four innermost circular waveguides correspond to each of the output ports of the switch. Switch Resonators allow a packet on an input port to be routed to any of the other output ports.

The local processor only has an input port because it receives packets via the buffers located at the other input ports. Each of the four circular waveguides in the centermost portion of the switch correspond to one of the four output ports. The North, South, East, and West input port waveguides connect to three of these output port circular waveguides through coupling resonators, and the Processor input port connects to all four. The blowup shows the Switch Resonators in the North Port and illustrates these connections where resonators enable the input waveguide to couple to the South, East and West ports. Similarly, its output waveguide couples to the portion of the switch corresponding to the North Port. Port buffers are located at the center of the router at each input port. A packet is buffered when it reaches its destination (final or interim; see Section 4.6) or if it is unable to win arbitration for its desired switch output port, causing it to block. In the latter case, no Switch Resonators will be set and the packet will be forced to enter the buffer.

The switch design eliminates the optical power loss associated with waveguide crossings through the use of two waveguide layers [Dokania and Apsel 2009]. Waveguide links connecting one router to another are implemented on a layer above the circular waveguides in the switch. Light couples between the two layers through the ring resonators in the switch and router input ports.

Unlike the Columbia approach [Shacham et al. 2007], our proposed optical switch has no electrical setup/teardown network. Rather, precomputed control bits for each router are optically transmitted in separate waveguides in parallel with the data, and these bits are used to implement simple dimension-order routing. Each packet consists of a single flit, which contains a full cache line (64 bytes) of Data, the Address, Operation Type, Error Detection/Correction, and miscellaneous bits. Router Control bits are also contained in the packet which are used at the source, intermediate and destination routers. Five waveguides contain the packet's payload, and a sixth waveguide holds the Router Control bits.

Theoretically, the Router Control could consist of 64 distinct routing groups, each of which corresponds to an individual node in the network. Prior to entering the network, a packet sets its Router Control by configuring only the routing groups corresponding to the switches it will traverse. All 64 possible routing groups each have seven different wavelengths corresponding to the four possible outputs plus Valid, Multicast, and Interim bits. In the simplest case all routing groups will be placed on a single waveguide such that every bit is implemented with a different wavelength. However, it is also

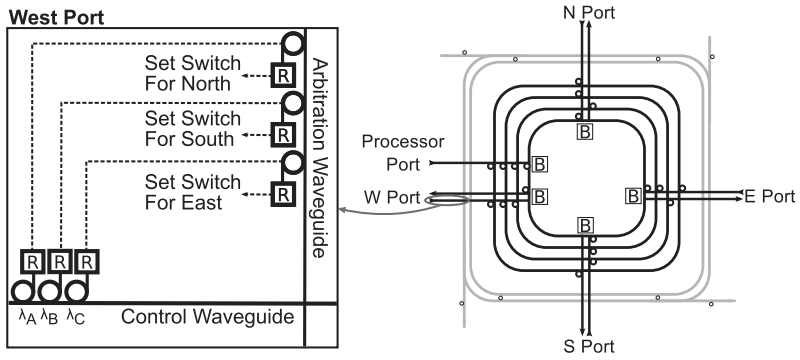


Fig. 3. Switch input ports receive control bits to set up the switch for proper routing. Three of the seven control bits are used for routing the packet to the proper output port. These control bits are received and used in switch arbitration.

possible to spread the groups across different waveguides to decrease wavelength usage. This is feasible because each router is statically configured to read its own routing group (i.e., proper wavelengths and waveguide) when a packet enters one of its input ports. The first four Router Control bits in a routing group map to the four possible outputs a packet can leave through in a router. If a packet enters through the North, East, South, or West ports, one of these bits represents the Local bit, which dictates whether the router should accept the packet for its local node. The Multicast and Interim bits will be discussed in Sections 4.5 and 4.6. The Valid bit is utilized in switch preconfiguration, which is introduced in Section 4.7.

In this work we implement an improvement over the use of routing groups corresponding to every router in the network in a packet's Router Control. We utilize routing groups that correspond to every input port in the network. Furthermore, if packets are routed deterministically, certain sets of input ports can share the same routing group since it is never the case that more than one of them can be used by a packet traveling to its destination. Compared to using per-hop routing groups, this permits reducing the number of required routing groups from 64 to 15, allowing us to fit all of a packet's Router Control on a single waveguide.

Consider a packet arriving at the West input port, as shown in Figure 3. The Control waveguide contains the seven control bits for this router in its associated control group, and the control bits for up to fourteen other routers depending on the distance between the packet's source and destination. Three of the control bits—East, North, and South—represent the desired route of the packet through the router. These bits are received and used to drive resonators connected to the Arbitration Waveguide where each of its resonators represent a different output port request. When a resonator is turned on, it generates a request for that output port. When arbitration has finished, the results are used to set the appropriate resonators in the switch.

It is important to note that a packet's payload data arrives in parallel with its control bits. Within each router, the control signals arbitrate for and set resonators to route the payload data through the switch. While this occurs, the payload data travels to the optical receiver just prior to the electrical buffering. If output port arbitration is won, the crossbar resonators are properly set and the payload data is routed around the circular waveguide and out the corresponding output port. This occurs through multiple switches within a given clock cycle. Thus, the control signals are on the critical path timing wise. If the packet doesn't win output port arbitration, none of the crossbar

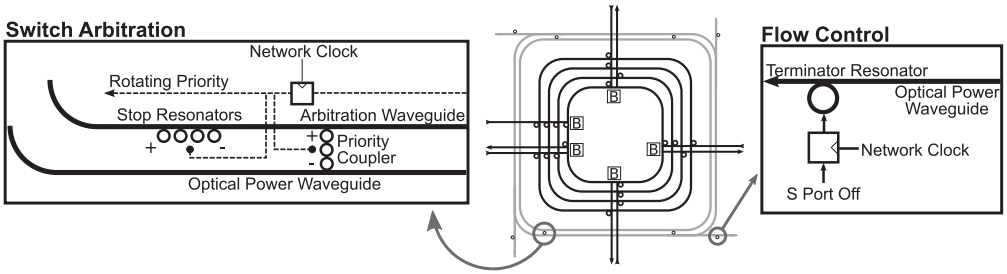


Fig. 4. Switch arbitration is composed of the two outermost circular waveguides in the optical router. An external laser source couples tokens into the Optical Power Waveguide at the four corners of the switch. Depending upon which priority coupler is activated, these tokens will proceed to couple into the Arbitration Waveguide at different points for use in switch arbitration. Stop Resonators are used to absorb the arbitration wavelengths that haven't been sinked by an input port. The Rotating Priority signal is passed in a rotating fashion to turn on a different Priority Coupler each cycle. Optical flow control utilizes the Optical Power Waveguide. If any of the token off signals are activated, Terminator Resonators prevent these tokens from being available for switch arbitration.

resonators are turned on and the packet stays at the receiver until the end of the cycle when it is latched into the electrical buffer.

All of a packet's routing, switch arbitration and switch setup operations are performed locally at each input port, which eliminates potentially high latency electrical operations associated with lengthy control signaling paths.

4.3. Switch Arbitration

Switch arbitration is enabled by the two outermost circular waveguides in the switch shown in Figure 4. Ring resonators (Priority Couplers) join the two waveguides at particular points in the loop, shown in the left blowup image. At each of the four corners of the switch light is coupled into the Optical Power Waveguide, which consists of four wavelengths (referred to as tokens), each corresponding to an output port in the switch. Every cycle only one Priority Coupler is activated by the Rotating Priority signal, allowing the light from the Optical Power Waveguide to couple into the inner Arbitration Waveguide. The switch arbitration priority changes every cycle as the Rotating Priority signal moves around the ring. The Stop Resonators prevent light from circulating around the Arbitration Waveguide more than once.

After a packet's control bits are translated to the electrical domain at a router's input port, they are used in switch arbitration. If a packet requests a particular output port, it will attempt to sink the wavelength associated with that output port's token. Light propagates in the counter clockwise direction in the Arbitration Waveguide, and input ports closest to the activated Priority Coupler in this direction have higher priority than others that are further away. When an input port arbitrates for an output port, it will sink the corresponding token by turning on the appropriate ring resonator along the Arbitration Waveguide such that any lower priority input ports no longer see that token wavelength. A packet on an input port may only exit an output port through the switch if it has its token. For example, consider the input port highlighted in Figure 3. If a packet enters this port the control wavelengths used for routing are received and used to drive an appropriate ring resonator on the Arbitration Waveguide. If the packet desires to be routed out the East Port, the third resonator from the top will be turned on. If the token for the East Port is available on the Arbitration Waveguide, it will be sinked off such that any lower priority input ports can no longer see it. Then it will be used to locally set the input port's Switch Resonators (see Figure 2) so that the packet can be properly routed to the East Port.

4.4. Electrical Buffering and Flow Control

Packets that do not couple into the switch waveguides continue to the buffers at the center of the switch. Here they are electrically received and latched into the input port's queue. In this study we implement on/off router flow control because it requires very little additional hardware complexity over what we have already discussed. The router flow control utilizes the switch Arbitration Waveguide through the Terminator Resonators, one of which is shown in the right blowup in Figure 4, which are located where light couples into the Optical Power Waveguide used for output port arbitration. Each Terminator Resonator corresponds to the wavelength of one of the output port tokens on the Arbitration Waveguide. If there are no free downstream buffer entries through an output port, the input ports should be prevented from sourcing a token for that output. A X Port Off signal, where X is North, South, East, or West, achieves this purpose. If a X Port Off signal is set, there will be no token corresponding to that output port available on the Arbitration Waveguide, forcing an incoming packet requiring that output to be buffered. Assuming that a downstream router can send an On or Off signal to an adjacent upstream router electrically in a single cycle, three buffers per input port are required to cover the roundtrip delay enabling full throughput. While a packet requires only a fraction of a cycle to travel across a network hop, a new packet will not utilize that same channel until the following network clock cycle.

At the beginning of each cycle, every input port buffer counts its number of free entries. If this number is one, an Off signal is sent upstream. On the following cycle, this signal latches into a register as shown in Figure 4 and turns off the appropriate token by turning on the corresponding Terminator Resonator. Similarly, when the number of free entries is two, no signal is sent and the Terminator Resonator is turned off the following cycle, allowing the token to flow.

Because of the delayed flow control signaling, it is possible for an input port to be transmitting a previously buffered packet and receiving a new packet in the same cycle. In order to avoid collisions between the two packets, the latter is bypassed to the center switch buffers via the Bypass Path shown in Figure 5. The Block Resonators, denoted in the diagram by resonators with a "B," prevent a failed packet transmission at the local input port from interfering with an incoming packet on the Bypass Path. When a packet is transmitted from an input port's queue, the Ongoing Transmission signal is activated, turning on the Block Resonators and Bypass Path. The Transmit Resonators are utilized to insert the packet transmission into the router just prior to the control logic used for routing, switch setup and arbitration. If the packet does not win its desired switch output port, it should not be rebuffered in the input port queue since it already exists at the head. Transmission failure is detected by noting the existence of a packet's Valid bit coupling through the Block Resonators. When this occurs, the input port knows to retransmit the packet at the head of the queue. Similarly, if the Valid bit is clear, it knows to pop off the packet at the head of its input port queue.

4.5. Multicast Operations

In a cache-coherent system, particular requests may be broadcast to every node. In the optical mesh topology that we evaluate in this study, a broadcast consists of multiple multicast packets. Multicast packets have the Multicast control bit set in the 7 bit router control group. For a 64-node system, the broadcasting node sends up to 16 multicast messages (eight if it is located on the top or bottom rows of the network).

For a given router, if the Multicast control bit is set, the Multicast Resonators are turned on as shown in Figure 5. The router then receives a portion of the power transmitted on the input lines through separate broadcast resonator/receivers via the Multicast Resonators. Since only a portion of the power is extracted, the packet continues through the selected output port to the next router in the absence of

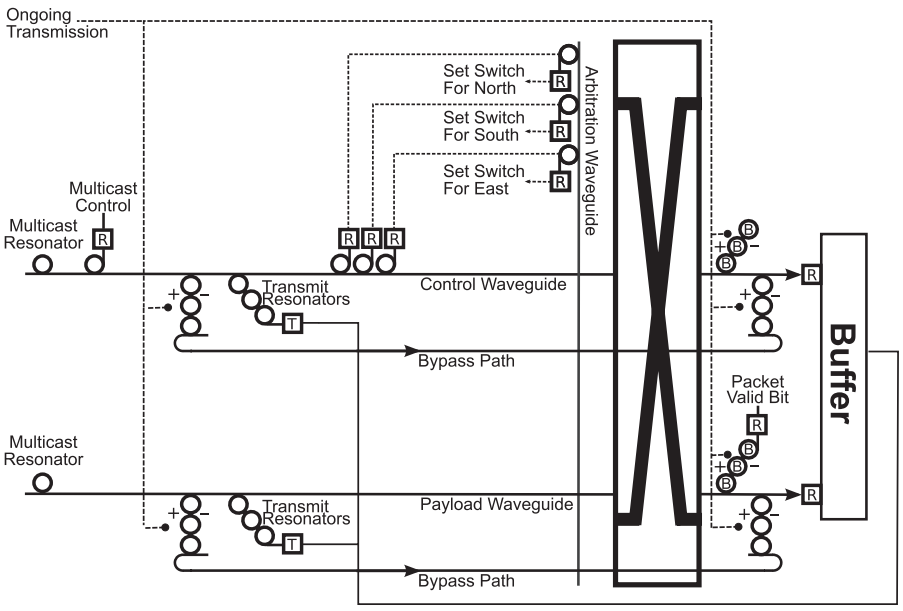


Fig. 5. Upon transmission in the network, a packet will utilize the Transmit Resonators to enter the router prior to the control logic. Any upstream packet that arrives on the same input port during a packet transmission must be buffered in order to avoid packet collisions. We do this through the Bypass Path and Block Resonators (designated by "B").

contention. The Multicast Resonators are placed prior to the Bypass and Transmit Resonators so that a packet does not perform unnecessary multicasts when blocked, buffered and retransmitted. One way to implement a Multicast Resonator is to vary its size such that its resonant frequencies are slightly shifted from the frequencies used to carry the network packets. This allows it to couple only a small percentage of the packet's power.

4.6. Interim Buffering

For large networks, all possible destinations may not be reachable in a single cycle. In these cases, the packet needs to be buffered at one or more interim nodes on its way to its final destination. We accomplish this using an Interim bit in every router's control group. When this bit it set, we force the packet to be buffered at that node. In the case that a packet can traverse eight hops in a network clock cycle, one way of implementing this is to set the Interim bit in every eighth router control group along its network path at the source node prior to its transmission. If a packet is prematurely buffered due to losing switch arbitration at a node, that node recalculates the Interim bits.

4.7. Switch Preconfiguration

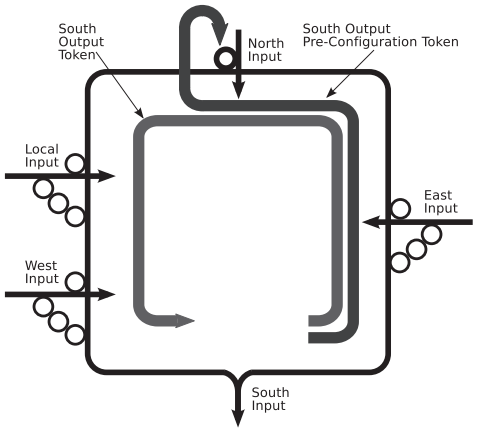
Because we implement dimension-order routing, a packet will spend most of its time traversing a router from the North port to the South port, East port to West port or vice versa. To minimize the per router hop delay, we implement a switch preconfiguration technique that statically joins the East/West and North/South router ports at the beginning of each network cycle prior to packet transmission. If an incoming packet enters an input port with a correctly configured output port, it continues through to the downstream router using a reduced latency path. Only when a packet desires an output port that differs from the straight output must it resort back to waiting for the control bits to properly set the switch as discussed in Section 4.3.

Input ports are statically configured to connect to the straight output ports through four additional tokens on the Arbitration Waveguide. We refer to these tokens as Pre-Configuration Tokens and they correspond to the North, East, South and West output ports. Thus the Arbitration Waveguide has four Pre-Configuration Tokens and the four Output Tokens described in Section 4.3. In Figure 6(a) the North port statically preconfigures itself at the beginning of the clock cycle by taking the Pre-Configuration Token for the South output from the Arbitration Waveguide. It uses this token to turn on the Switch Resonators for connecting to the South output. The South Output Token remains on the Arbitration Waveguide. Following switch preconfiguration a packet may enter the North input requiring the South output as shown in Figure 6(b). Because the switch was previously set up to make this connection, the packet can traverse the router with a reduced delay path. Roundabout waveguides allow the packet to bypass the Switch Resonators of the other input ports. A roundabout waveguide in combination with the switch waveguide that it is attached to functions as an asymmetric y-branch for variable power splitting [Sakat et al. 2002; Lin et al. 1999]. Any light that is traveling through the switch that does not belong to the input port corresponding to the roundabout will couple into it entirely, allowing it to bypass that input port's switch resonators. Additionally, light from an input port will not couple into its own roundabout because of the y-branch functionality. For simplicity, the following discussion on switch preconfiguration will only refer to the input ports connected to the South output.

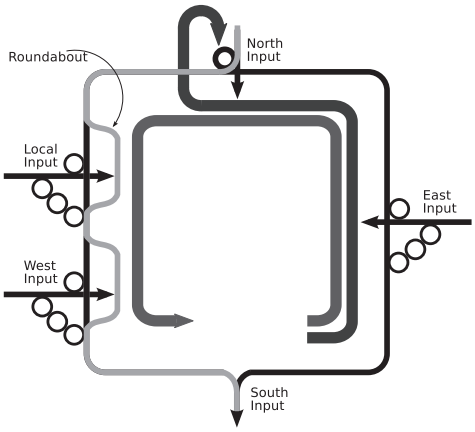
Switch preconfiguration still respects the rotating priority arbitration scheme introduced in Section 4.3. Any packet that enters the switch on the Local, West, or East inputs and requires the South output will attempt to take both the South Pre-Configuration Token and the South Output Token. Packets on input ports with lower arbitration priority than the North can only access the South Output Token. This is shown in Figure 6(c) where the lower priority West input routes to the South using the Output Token to turn on one of its two sets of Switch Resonators. One set is turned on by the South Pre-Configuration Token and the other by the South Output Token, where the former takes precedence. Thus because the West input was only able to take the Output Token for the South, it traverses the switch in the direction that forces it to pass by the North input. If in the same cycle a packet on the North enters the router and simultaneously requires the South output, it will take the South Output Token away from the West input and turn on the Pre-Configuration Block Resonator. This is denoted in Figure 6(d) by the resonator with "PB." When this resonator is turned on, the packet from the West input is blocked from leaving through the South, allowing the North packet to traverse the switch without having to wait for it to buffer. Shortly after, the West input will be forced to turn off its Switch Resonators since it no longer has the the Output Token and buffer as shown in Figure 6(e). However, the incoming packet on the North does not have to wait for this to occur before leaving the router.

In Figure 6(f) the East input has higher arbitration priority than the North input, allowing an incoming packet there to take both tokens for the South output. In this case both sets of Switch Resonators are turned on, but precedence is given to the set turned on by the Pre-Configuration Token, which routes the packet in the crossbar away from the North input. This is because if in the same cycle a packet on the North port enters the router and also wants to leave through the South, it will turn on its Pre-Configuration Block Resonator regardless of whether it still has access to the switch. However, because the North packet no longer has its Switch Resonators turned on, it will be buffered as shown in the diagram.

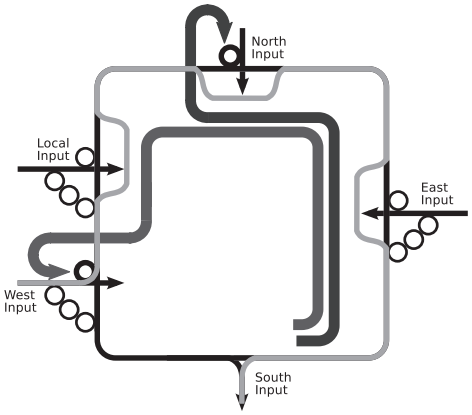
If an incoming packet on the North port requires other than the preconfigured South port, the packet must turn on the appropriate resonators in the Arbitration Waveguide, attempting to take both the Pre-Configuration and Output Token for the desired output.



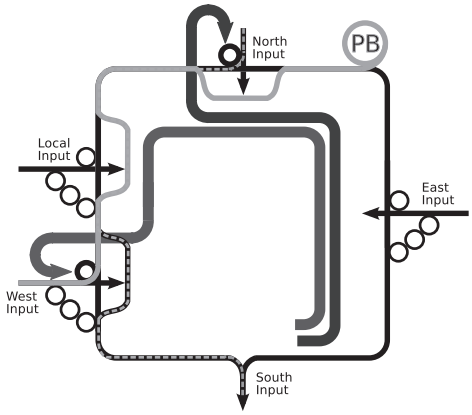
(a) North input is preconfigured for South output.



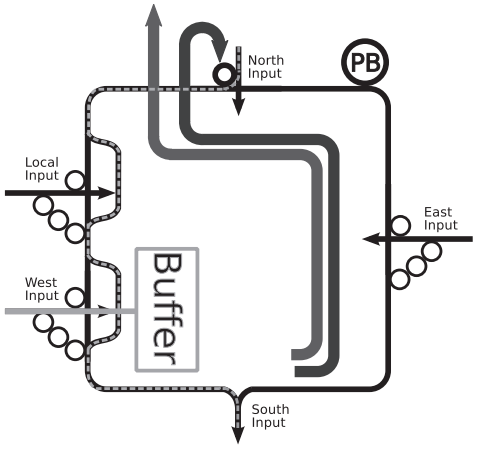
(b) North packet uses preconfigured route.



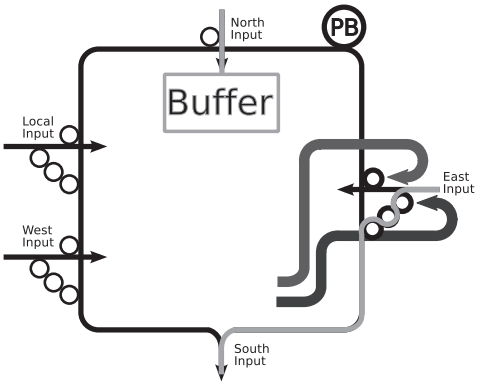
(c) Lower priority West packet uses South output.



(d) West packet blocked by North packet.



(e) West packet loses Output Token and buffers.



(f) High priority East packet uses South output.

Fig. 6. East, West, North, and South inputs are statically preconfigured to connect to straight path output ports. For clarity, only the ports connecting to the South output are shown.

Table I. Predicted Optical Component Delay Values for 16nm

Component	Experimental delay (ps)
Driver + Resonator	11.3 ps
Optical receiver	1.32 ps
Optical signal propagation	6 ps/mm [Gondarenko et al. 2009]

Thus the router delay for a packet that enters a port with an incorrectly preconfigured path, or a packet that enters through the Local input port (which does not have a preconfigured route) is the same as when Switch Pre-Configuration is not supported.

The addition of Switch Pre-Configuration requires the flow control implementation to be slightly modified. When a downstream buffer is full, the flow control signal that propagates back must turn on Terminator Resonators for both the Output Token and the Pre-Configuration Token.

5. OPTICAL ROUTER DESIGN ANALYSIS

5.1. Critical Delay

The critical delay timing components of our proposed optical switch architecture can be divided into three broad categories. The first category is denoted as Router Setup, which is composed of switch tasks that are completed prior to packets transmitting into the network. These tasks consist of setting up the optical switch arbitration including turning on the appropriate Priority Coupler, Stop Resonators and propagating the output tokens around the Arbitration Waveguide. This step also involves turning on the Transmit Resonators and associated Bypass Resonators and Block Resonators. Lastly, flow control signals from downstream routers which propagated the previous cycle are used to set the proper Terminator Resonators on the Optical Power Waveguide. In parallel with router setup, if supported, switch preconfiguration will begin to statically configure the network switches.

The second timing category is denoted as Router Traversal and consists of two possible delay paths through a network router. The first path requires the packet to wait for control bit translation, switch arbitration and setup prior to entering the crossbar. This occurs when, a) a packet enters through the Local input, which has no preconfigured route, b) a packet makes a turn, or c) preconfiguration is not supported. The second type of path occurs when switch preconfiguration is supported and matches the desired output of a packet. Here the optical packet continues through the router with a minimally impeded delay.

The last timing category, denoted as Cycle Termination, occurs at the end of a clock cycle when a packet enters into an input port and uses its Interim control bits to buffer. This consists of receiving the Interim control signal, turning on the Bypass Path and buffering the packet.

In parallel with network packet transmission, each input port buffer performs an appropriate flow control action. This involves determining the number of free buffer entries and sending an off signal to an upstream router if necessary.

The individual delay parameters for the optical components used in our critical delay analysis are found in Table I. The Driver + Resonator and Optical receiver delays are scaled values of Kirman et al. [2006]. We determine that our switch preconfiguration scheme allows a packet to traverse eight hops in a single 4GHz network cycle, versus only four hops with no preconfiguration.

5.2. Area

The area of the optical components in our proposed router should not exceed the area of the electrical components in a network node, otherwise the latter will need to

artificially increase in size to line up the related components. Moreover, the electrical components of the router, such as the resonator drivers and receiver amplifiers, should only marginally increase the area of the processor die.

To estimate the area of the processor die, we adopted the methodology of Kumar et al. [2006] for 16nm technology. For a single processor core with 64 KB L1 caches, a 2MB L2 cache, and Memory Controller the total area is approximately 3.5mm^2. For two cores and four cores sharing an L2 cache, the area is approximately 4.5mm^2 and 6.5mm^2, respectively. The area of the optical components of our proposed router consume approximately 3.1mm^2 under the assumption that a router's datapath uses six waveguides to route a packet, allowing it to be deposited above the processor without the need to grow its area. The electrical components of the optical network which facilitate the communication between the electrical and optical domains (i.e., receiver amplifiers and transmitter driver circuitry), consume approximately .12mm^2 per router on the electrical die. This represents a 3% area overhead over a single processing core.

5.3. Optical Power

In this work we assume that a laser externally supplies light to the on-chip interconnect through vertical coupling which incurs a 3 dB loss [Kirman et al. 2006]. Additionally, laser efficiency is assumed to be 50%, which adds another 3 dB of optical loss. Propagation losses in the silicon nitride waveguides are .1 dB/cm [Gondarenko et al. 2009]. The polycrystalline resonators consist of two different loss components. For light that couples into a resonator we assume a 1 dB loss. When light passes a resonator without coupling into it, a .1 dB loss occurs [Kirman et al. 2006]. We utilize ring modulators which incur a 1 dB loss and modulate wavelengths at a rate of 4 Gb/s [Preston et al. 2008]. Polycrystalline germanium detectors each have a responsivity of 1 A/W [Preston et al. 2010]. We calculate the worst case laser power requirements of the network. This situation occurs when 16 packets are sent the maximum of 8 hops in a single network cycle, broadcasting a portion of their power along the way. The laser is always on and thus contributes to the static power consumption of the network, albeit externally to the chip. We estimate that the chip will require 35W of optical power to handle this worst case scenario.

6. EVALUATION METHODOLOGY

To evaluate our proposed optical network, we developed a cycle-accurate network packet simulator that models components down to the flit-level. The simulator generates traffic based on a set of input traces that designate per node packet injections. In order to do a power comparison with the electrical baseline, we model dynamic power consumption and static leakage power in a manner similar to [Kirman et al. 2006]. The optical ring resonators are very sensitive to fluctations in die temperature. In order to account for these fluctations we also model the power consumed by resistive in-plane heaters and their associated feedback circuits.

We evaluate the electrical baseline network using a modified version of Booksim [Dally and Towles 2007] augmented with dynamic and static leakage power models. The models use CACTI for buffers, and the methodology of Balfour and Dally [2008] for all other components. We also implemented Virtual Circuit Tree Multicasting [Jerger et al. 2008] to perform packet broadcasts. Finally, we changed Booksim to input the same trace files used for our optical simulator.

The electrical baseline is an aggressive router optimized for both latency and bandwidth. The router assumes a virtual-channel architecture with the parameters shown in Table II. In order to perform a fair performance comparison with our optical configurations, we assume both low latency and high saturation bandwidth for the electrical network. We reduce serialization latency by using a packet size of one flit, the same as

Table II. Baseline Electrical Router Parameters

Flits per Packet	1 (80 bytes)
Routing Function	Dimension-Order
Number of VCs per Port	4
Number of Entries per VC	1
Wait for Tail Credit	YES
VC_Allocator	ISLIP [McKeown 1999]
SW_Allocator	ISLIP [McKeown 1999]
Total Router Delay	2 cycles

Table III. Memory Parameters

Simulated Cache Sizes	32KB L1I&L1D, 256KB L2
Actual Cache Sizes	64KB L1I&L1D, 2MB L2
Cache Associativity	4 Way L1,16 Way L2
Block Size	32B L1, 64B L2
Memory Latency	80 Cycles

in our proposed architecture. Doing so also gives no bandwidth density advantage to the optical network. We further assume that pipeline speculation and route-lookahead [Peh and Dally 2001] reduce the per hop router latency of the baseline electrical router to 2 cycles for every flit. The bisection bandwidth of the electrical and optical systems are matched at 4TB per second.

We evaluate SPLASH2 benchmarks and synthetic traffic workloads. By varying the injection rates of the synthetic benchmarks, we obtain saturation bandwidth and average packet latencies. We created SPLASH2 traces using the SESC simulator [Renau et al. 2005]. The modeled system consists of 64 cores with private L1 and L2 caches. Each core is 4-way out-of-order and has the cache and memory parameters shown in Table III. As is typical when using SPLASH2 for network studies, the cache sizes are reduced to obtain sufficient network traffic. Finally, we assume a 16nm technology node operating at a 4GHz processor and network clock with a supply voltage of 1.0V.

7. RESULTS

The results that we present in this section compare the electrical baseline, denoted as Electrical, with three optical configurations. Using the delay calculations in Section 5.1 we determine that our proposed optical routers can achieve four hops in a network cycle when no switch preconfiguration is used. This network configuration is called No Preconfig in our results. When preconfiguration is used, a packet can traverse up to eight hops in a cycle and is denoted as Preconfig. We also present a set of upper bound results where we assume that a packet can optically traverse the entire length of the network, reaching its destination in a single clock cycle if it is not blocked at an interim node. These results are denoted as Perfect.

7.1. Microbenchmark Analysis

In this section we motivate the use of optical switch preconfiguration in the context of our analyzed 64 processor mesh network. We provide a small case study in which only four memory controllers exist in the system, each of which is connected to one of the routers at the four corners of the network grid. Thus all traffic is forced to these four corners of the die. Main memory is distributed and cache line interleaved between the four access points.

We analyzed the Splash benchmarks [Woo et al. 1995] and determined that on average each program's instruction set consists of 30% loads and stores. In Figure 7 we present results for a synthetic benchmark that models multiprocessor network traffic relative to the No Preconfig architecture. Each processor injects into the network

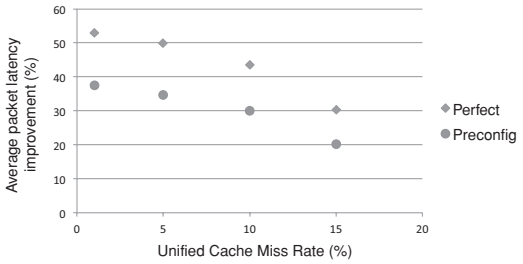
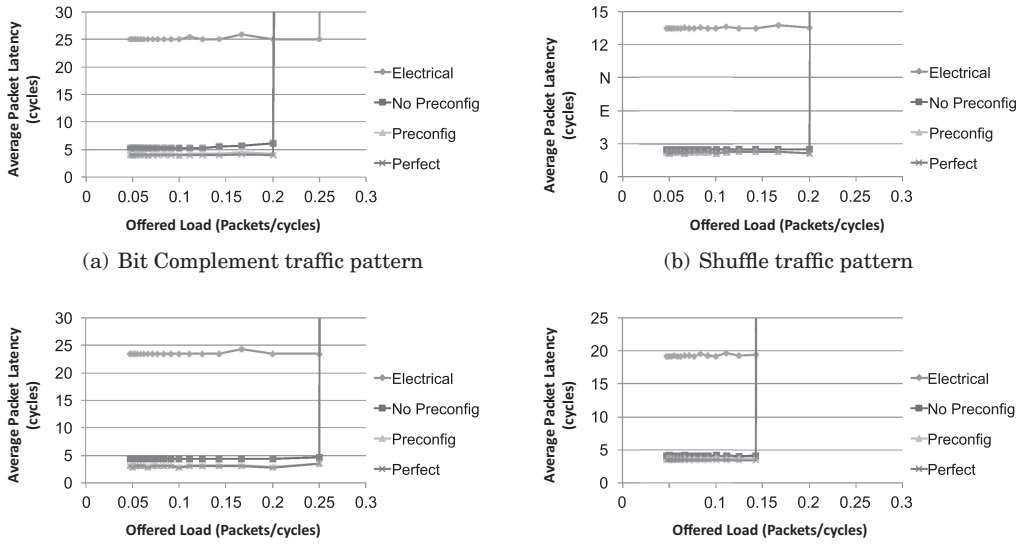


Fig. 7. Synthetic benchmark representing off-chip memory accesses through four on-chip memory controllers. The y-axis represents the average packet latency improvement when optical switch preconfiguration is used relative to the switch with no preconfiguration. The x-axis varies the unified cache miss rate of each processing node, where a higher miss rate results in more network injections.



(a) Bit Complement traffic pattern

(b) Shuffle traffic pattern

(c) Tornado traffic pattern

(d) Transpose traffic pattern

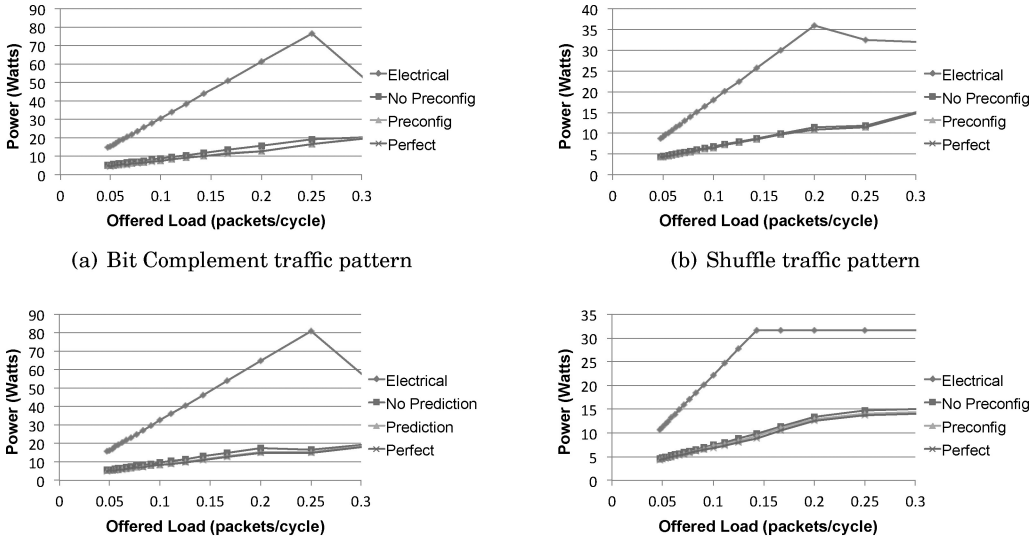
Fig. 8. Synthetic benchmark performance results. The y-axis represents the average packet latency under varying offered loads (packets/cycle).

according to a probability that on a given cycle the processor will attempt to access its memory hierarchy and also miss in the last level (L2) cache. If this occurs, a network packet request is created with an equal probability of traveling to one of the four memory controllers. We show results for different unified cache miss rates ranging from 1% to 15%.

For unified cache miss rates below 15%, the Perfect configuration achieves 30% to 53% lower average packet latencies compared to the optical switch with no preconfiguration. Across the four different unified cache miss rates, Perfect achieves 40% better performance on average and Preconfig performs 30% better than No Preconfig. As the unified cache miss rate is increased to 15%, more packets are injected into the network. At this point the network performance benefits of using switch preconfiguration are lower, yet still significant at 20%, as the network becomes increasingly congested.

7.2. Synthetic Benchmark Results

We analyze four synthetic benchmarks and report latency versus offered load for each network configuration, as shown in Figure 8. The optical configurations are able to



(a) Bit Complement traffic pattern

(b) Shuffle traffic pattern

(c) Tornado traffic pattern

(d) Transpose traffic pattern

Fig. 9. Synthetic benchmark power results. The y-axis represents the average network power consumption in Watts under varying offered loads (packets/cycle).

take advantage of the high-speed optical routers to achieve more hops in a network cycle than the electrical baseline. On average the No Preconfig architecture requires 16 fewer cycles than the electrical baseline network to route packets in the network prior to saturation. Perfect has on average 21% lower latency than No Preconfig across all the synthetic loads, and 5% lower than Preconfig. Switch preconfiguration can be unnecessary in some of the synthetic benchmarks because packets do not need to traverse a distance greater than four hops to reach their destination.

The saturation characteristics of each optical network configuration are well matched to the electrical baseline. The electrical network only saturates at a slighly higher offered load in Bit Complement. While rotating priority switch arbitration can guarantee forward progress to all nodes in the network, it does not offer local fairness like ISLIP allocation in the electrical baseline. This is because the switch tokens propagate into the arbitration waveguide at different points depending upon which Priority Coupler is activated. In a radix 5 router, there are five Priority Couplers. Because the token propagates in the counter-clockwise direction, the two router ports to the left of any port will have higher priority in four of the five possible arbitration waveguide priority couplings. Future work will examine possible ways to achieve local fairness in switch allocation. Additionally, the virtual channels in the electrical baseline, which are not present in our optical router, enable "turning lanes" which help increase the network saturation point.

The power results for the synthetic workloads are shown in Figure 9. The optical configurations are far more efficient than the electrical baseline network across all offered loads. The power of the electrical baseline in Shuffle, Bit Complement and Tornado dips slightly after saturation due to network instability. This is not present in the optical configurations. Ideally, as the source injection loads increase beyond the saturation point, the actual bandwidth routed within the network should remain steady. In three of the four synthetic benchmarks the electrical switch allocation algorithm is unable to provide global resource fairness at high network loading, forcing some source, destination pairs to inject at reduced bandwidths following saturation

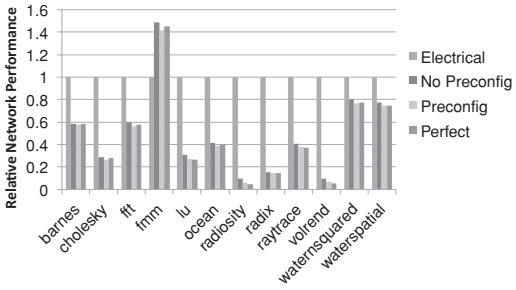


Fig. 10. Splash Performance Comparison.

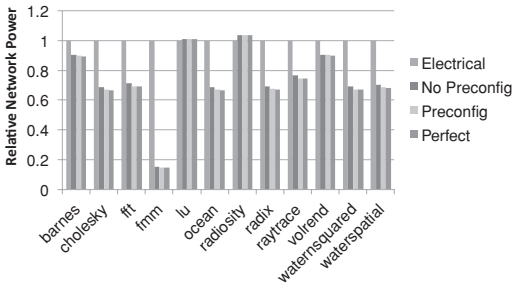


Fig. 11. Splash Power Comparison.

[Dally and Towles 2007]. This is not the case in the rotating priority allocation used in the optical networks. In all of the synthetic benchmarks that were analyzed, the reduced local fairness of rotating priority enables improved global fairness. This is because a packet can potentially traverse many hops to reach a shared channel without consistently losing switch arbitration in an upstream router.

7.3. Splash Benchmark Results

Figure 10 shows relative performance results for the Splash benchmarks. The optical configurations achieve on average a 2X speedup in average packet latencies across the applications. FMM is the only benchmark where the electrical baseline achieves better performance. As was the case in the Bit Complement synthetic benchmark, the electrical network benefits from virtual channel paths which reduces packet blocking, and thus increases saturation bandwidth over the optical architectures.

In seven of the benchmarks the Preconfig configuration achieves marginally better performance than the network with Perfect routers. This is because of the slight unfairness inherent in our rotating priority switch arbitration scheme. When a packet can traverse the entire length of the network, as in Perfect, it may create a situation where channel bandwidth is unevenly shared in some of the programs. In the event that the number of hops a packet can reach in a cycle is reduced, such as in Preconfig, other packets will be able to utilize the freed bandwidth for themselves. However, there is an optimal point where performance begins to decrease as the number of hops a packet can reach in a cycle is further reduced. This is demonstrated by the results for No Preconfig.

On-chip power consumption results for the Splash benchmarks are shown in Figure 11. All of the Splash benchmarks except for FMM typically have low injection rates and thus the static power overhead of the resonator heaters in the optical network becomes more noticeable. This is also the case in the synthetic benchmarks

for very low injection rates, where the power consumption of the electrical and optical network are similar. As injection rates climb, however, the power consumption benefits of the optical network increase as the static overhead of resonator heaters decreases. Two benchmarks in particular, Lu and Radiosity, exhibit situations where the optical power consumption slightly surpasses the electrical network due to the very low injection activity throughout the application. Overall, however, the optical networks achieves 30% lower power consumption on average across all applications.

8. CONCLUSIONS

In this article we introduce a novel optical router architecture that exploits the low latency of nanophotonics to permit multihop packet traversal in a network clock cycle. On the set of Splash bechmarks we examined, our optical configurations achieve 2X better network performance on average while consuming 30% less power.

Future work will examine a switch allocation algorithm that provides increased local fairness. We would also like to integrate virtual channels into our optical router architecture and explore the trade-offs in varying the radix for use in different network topologies.

REFERENCES

ALMEIDA, V. R., BARRIOS, C. A., PANEPUCCI, R. R., LIPSON, M., FOSTER, M. A., OUZOUNOV, D. G., AND GAETA, A. L. 2004. All-optical switching on a silicon chip. *Opt. Lett. 29,* 24.

BALFOUR, J. AND DALLY, W. 2008. Design tradeoffs for tiled cmp on-chip networks. In *Proceedings of the International Conference on Supercomputing*.

BARWICZ, T., POPOVI, M., RAKICH, P., WATTS, W., HAUS, H., IPPEN, E., AND SMITH, H. 2004. Microring-resonator-based add-drop filters in sin: fabrication and analysis. *Opt. Express 12,* 7.

BATTESTILLI, T. AND PERROS, H. 2003. An introduction to optical burst switching. *IEEE Opt. Comm. 41,* 8.

BOGAERTS, W., DUMON, P., THOURHOUT, D. V., AND BAETS, R. 2007. Low-loss, low-cross-talk crossings for silicon-on-insulator nanophotonic waveguides. *Opt. Lett. 32,* 19.

CHEN, G., CHEN, H., HAURYLAU, M., NELSON, N., FAUCHET, P., FRIEDMAN, E., AND ALBONESI, D. 2005. Predictions of cmos compatible on-chip optical interconnect. In *Proceedings of the International Workshop on System Level Interconnect*.

CHEN, Y., QIAO, C., AND YU, X. 2004. Optical burst switching: A new area in optical networking research. *IEEE Netw. Mag. 18,* 3.

CIANCHETTI, M., KEREKES, J., AND ALBONESI, D. 2009. Phastlane: A rapid transit optical routing network. In *Proceedings of the International Symposium on Computer Architecture*.

DALLY, W. AND TOWLES, B. 2007. *Principles and Practices of Interconnection Networks*. Morgan Kaufmann.

DOKANIA, R. AND APSEL, A. 2009. Analysis of challenges for on-chip optical interconnects. In *Proceedings of the Great Lakes Symposium on VLSI*.

GONDARENKO, A., LEVY, J., AND LIPSON, M. 2009. High confinement micron-scale silicon nitride high q ring resonator. *Opt. Express 17,* 14.

ITRS. 2008. International Technology Roadmap for Semiconductors (ITRS). The ITRS Technology Workshop Groups, http://public.itrs.net.

JERGER, N. E., PEH, L.-S., AND LIPASTI, M. 2008. Virtual circuit tree multicasting: A case for on-chip hardware multicast support. In *Proceedings of the International Symposium on Computer Architecture*.

JOSHI, A., BATTEN, C., KWON, Y., BEAMER, S., SHAMIM, I., ASANOVIC, K., AND STOJANOVIC, V. 2009. Silicon-photonic clos networks for global on-chip communication. *Opt. Lett. 29,* 24.

KIRMAN, N., KIRMAN, M., DOKANIA, R., MARTINEZ, J., APSEL, A., WATKINS, M., AND ALBONESI, D. 2006. Leveraging optical technology in future bus-based chip multiprocessors. In *Proceedings of the International Symposium on Microarchitecture*.

KIRMAN, N. AND MARTINEZ, J. 2010. An efficient all-optical on-chip interconnect based on oblivious routing. In *Proceedings of the Conference on Architectural Support for Programming Languages and Operating Systems*.

KUMAR, A., PEH, L.-S., KUNDU, P., AND JHA, N. 2007. Express virtual channels: Towards the ideal interconnection fabric. In *Proceedings of the International Symposium on Computer Architecture*.

KUMAR, R., TULLSEN, D., AND JOUPPI, N. 2006. Core architecture optimization for heterogeneous chip multiprocessors. In *Proceedings of the International Symposium on Parallel Architectures and Compilation Techniques*.

LIN, H., SU, J., CHENG, R., AND WANG, W. 1999. Novel optical single-mode asymmetric y-branches for variable power splitting. *IEEE J. Quantum Electron. 35,* 7.

LIU, C., TAM, H., WAI, P., AND PUN, E. 2004. Low-loss waveguide crossing using a multimode interference structure. *Opt. Comm. 241,* 1–3.

MCKEOWN, N. 1999. The islip scheduling algorithm for input-queued switches. *ACM Trans. Net. 7,* 2.

MELCHIORRI, M., DALDOSSO, N., SBRANA, F., AND PAVESI, L. 2005. Propagation losses of silicon nitride waveguides in the near-infrared range. *Appl. Phys. Lett. 86,* 12.

MILLER, D. 2000. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE 88,* 6.

PANICCIA, M., KRUTUL, V., JONES, R., COHEN, O., BOWERS, J., FANG, A., AND PARK, H. 2006. A hybrid silicon laser. White Paper, Intel Corporation.

PARK, H., HAO KUO, Y., FANG, A. W., JONES, R., COHEN, O., PANICCIA, M. J., AND BOWERS, J. E. 2007. A hybrid algainas-silicon evanescent preamplifier and photodetector. *Opt. Express 15,* 21.

PEH, L.-S. AND DALLY, W. 2001. A delay model and speculative architecture for pipelined routers. In *Proceedings of the International Symposium on High-Performance Computer Architecture*.

PRESTON, K., DONG, P., SCHMIDT, B., AND LIPSON, M. 2008. High-speed all-optical modulation using polycrystalline silicon microring resonators. *Appl. Phys. Lett. 92,* 15.

PRESTON, K., SCHMIDT, B., AND LIPSON, M. 2007. Polysilicon photonic resonators for large-scale 3d integration of optical networks. *Opt. Express 15,* 25.

PRESTON, K., ZHANG, M., AND LIPSON, M. 2010. Waveguide-integrated photodiode in deposited silicon. In *Proceedings of the Conference on Integrated Photonics Research, Silicon and Nanophotonics*.

RENAU, J., FRAGUELA, B., TUCK, J., LIU, W., PRVULOVIC, M., CEZE, L., SARANGI, S., SACK, P., STRAUSS, K., AND MONTESINOS, P. 2005. Sesc simulator. http://sesc.sourceforge.net.

SAKAT, A., FUKAZAWA, T., AND BABA, T. 2002. Low loss ultra-small branches in a silicon photonic wire waveguide. *IEEE Trans. Electron. E85C,* 4.

SHACHAM, A., BERGMAN, K., AND CARLONI, L. 2007. On the design of a photonic network-on-chip. In *Proceedings of the International Symposium on Networks-on-Chip*.

SHAW, M., GUO, J., VAWTER, G., HABERMEHL, S., AND SULLIVAN, C. 2005. Fabrication techniques for low loss silicon nitride waveguides. In *Proc. SPIE Micromachining Technology for Micro-Optics and Nano-Optics III*.

TATUM, J. 2001. Vcsels for 10 gb/s optical interconnects. *Proceedings of the IEEE Symposium on Emerging Technologies. Broadband Communications for the Internet Era Symposium Digest*.

VANTREASE, D., SCHREIBER, R., MONCHIERO, M., MCLAREN, M., JOUPPI, N., FIORENTINO, M., DAVID, A., BINKERT, N., BEAUSOLEIL, R., AND AHN, J. 2008. Corona: System implications of emerging nanophotonic technology. In *Proceedings of the International Symposium on Computer Architecture*.

WOO, S. C., OHARA, M., TORRIE, E., SINGH, J. P., AND GUPTA, A. 1995. The splash-2 programs: Characterization and methodological considerations. In *Proceedings of the International Symposium on Computer Architecture*.

XIA, F., SEKARIC, L., AND VLASOV, Y. 2006. Ultracompact optical buffers on a silicon chip. *Nature Photon. 1,* 1.

XU, Q. AND LIPSON, M. 2007. All-optical logic based on silicon micro-ring resonators. *Opt. Express 15,* 3.

XU, Q., MANIPATRUNI, S., SCHMIDT, B., SHAKYA, K., AND LIPSON, M. 2007. 12.5 gbit/s carrier-injection-based silicon micro-ring silicon modulators. *Opt. Express 15,* 2.

YANAGASE, Y., SUZUKI, S., KOKUBUN, Y., AND CHU, S. 2002. Box-like filter response and expansion of fsr by a vertically triple coupled microring resonator filter. *J. Lightwave Techn. 20,* 8.

YOUNG, I., MOHAMMED, E., LIAO, J., KERN, A., PALERMO, S., BLOCK, B., RESHOTKO, M., AND CHANG, P. 2010. Optical i/o technology for tera-scale computing. *IEEE J. Solid-State Circ. 45,* 1.

Received July 2010; revised November 2010, February 2011; accepted March 2011