Analysis of Photonic Networks for a Chip Multiprocessor Using Scientific Applications

Gilbert Hendry[†], Shoaib Kamil[‡⋆], Aleksandr Biberman[†], Johnnie Chan[†], Benjamin G. Lee[†], Marghoob Mohiyuddin[‡⋆], Ankit Jain[‡], Keren Bergman[†], Luca P. Carloni[γ], John Kubiatowicz[‡], Leonid Oliker[⋆], John Shalf[⋆]

[†] *Lightwave Research Laboratory, Columbia University, New York, NY 10027*
[γ] *Computer Science Department, Columbia University, New York, NY 10027*
[⋆] *CRD/NERSC, Lawrence Berkeley National Laboratory, Berkeley, CA 94720*
[‡] *Computer Science Department, University of California, Berkeley, CA 94720*

Abstract

As multiprocessors scale to unprecedented numbers of cores in order to sustain performance growth, it is vital that these gains are not nullified by high energy consumption from inter-core communication. With recent advances in 3D Integration CMOS technology, the possibility for realizing hybrid photonic-electronic networks-on-chip warrants investigating real application traces on functionally comparable photonic and electronic network designs. We present a comparative analysis using both synthetic benchmarks as well as real applications, run through detailed cycle accurate models implemented under the OMNeT++ discrete event simulation environment. Results show that when utilizing standard process-to-processor mapping methods, this hybrid network can achieve 75× improvement in energy efficiency for synthetic benchmarks and up to 37× improvement for real scientific applications, defined as network performance per energy spent, over an electronic mesh for large messages across a variety of communication patterns.

1 Introduction

The microprocessor industry is set to double the number of cores per chip every 18 months – leading to chips containing hundreds of processor cores in the next few years. This path has been set by a number of conspiring forces, including complexity of logic design and verification, limits to instruction level parallelism and – most importantly – constraints on power dissipation. In this brave new world of ubiquitous chip multiprocessing (CMP), the on-chip interconnect will be a critical component to achieving good parallel performance. Unfortunately, a poorly designed network could easily consume significant power, thereby nullifying the advantages of chip multiprocessing.

Consequently, we must find communication architectures that can somehow maintain performance growth under a fixed power budget. Current processor-manufacturing roadmaps point to simple mesh or torus networks-on-chip (NoC) via electrical routers as the medium-term solution; however, previous work [1] has shown that such architectures may not be best-suited for balancing performance and energy usage. In this paper, we investigate a promising alternative to electrical NoCs, namely architectures that exploit optics for some or all inter-processor communications.

According to the International Technology Roadmap for Semiconductors [10], three-dimensional chip stacking for three-dimensional integration (3DI) is a key focus area for improving latency and power dissipation, as well as for providing functionally diverse chip assemblies. Recent advances in 3DI CMOS technology [3] have paved the way for the integration of silicon-based nanophotonic devices with conventional CMOS electronics, with the premise of realizing hybrid photonic/electronic NoCs [17]. High density through-silicon-vias (TSVs), the critical enabling technology for 3DI, electrically connect wafer layers. One of the fundamental assumptions of this work is that 3D integrated chips will play an important role as the interconnect plane for future chip multiprocessors, whether the NoC is electrical or photonic, and that the TSVs have a minimal impact on the power dissipation for these chip implementations.

To evaluate the tradeoffs between the electrical and photonic network designs, we conduct extensive cycle-accurate simulations using custom software within the OMNeT++ framework [19]. This work differs from previous efforts through the use of a comprehensive event-driven simulation allowing us to model the low-level electronic and photonic details of the evaluated interconnect configurations. The modeling detail enables us to analyze the energy, latency, and physical performance of the devices. In addition to standard synthetic traffic models, our study utilizes traces of real parallel scientific applications to determine the potential benefits of the hybrid network for Single Program

978-1-4244-4143-3/09/$25.00

Multiple Data (SPMD) style algorithms.

The simulation environment is used to analyze interconnection networks of various types and configurations for performance and energy consumption. Reported metrics include the execution time of the benchmark/application, the total energy consumed therein, and the energy efficiency, a metric which emphasizes the network performance gained with each unit of energy spent. We simulate the performance of electronic mesh and torus topologies along with the photonic NoC studied in [15], known as a blocking torus (which we refer to as a *photonic torus*). In this photonic NoC, a photonic network and an electronic control network coordinate to provide the system with high bandwidth communications. The simulations show that the photonic interconnects studied here offer excellent power-efficiency for large messages, but are less advantageous for carrying small messages. We present a detailed set of results that show how different application characteristics can affect the overall performance of the network in ways that are not readily apparent in higher level analysis.

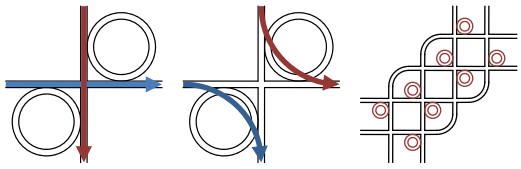
2 Related Work

Prior related works have made significant gains in the area of on-chip optical interconnects. Petracca *et al.* investigated Cooley-Tukey FFT traffic patterns on different photonic topologies in [15]. The photonic NoC is described as an electronic control network augmented with a photonic network made up of silicon waveguides and *photonic switching elements* (PSEs). Each PSE, shown in Figure 1, is composed of silicon micro-ring resonators that deflect light when polarized. These building blocks are extended to create a broadband circuit-switched 2D torus topology for on-chip communication.

Novel wavelength-routed architectures have also been proposed both for inter-core communications [18] and for off-chip communications [2]. These networks take advantage of wavelength-division multiplexing (WDM) to dedicate wavelengths to destinations in the network. Lower level modeling was performed in [5, 14], which is a good step towards achieving a comprehensive analysis of an architecture, but it has yet to be seen how these networks compare to other competing systems under real workloads.

For electronic CMPs, Dally *et al.* [1] compared several possible NoC topologies using detailed timing, area, and energy models for the network components. Of the explored networks, the best in terms of energy and communication time was a *Concentrated Mesh*, a type of mesh topology that uses larger-radix routers to cluster four processors at each mesh node and contains express channels around the perimeter of the network.

Other work proposing a hybrid interconnection network for multiple processor systems [11] characterized the inter-chip communication requirements for full scientific applications using similar measurement tools. The study found that fully connected network topologies are overprovisioned for most applications and their size grows exponentially with system concurrency. However, mapping application communication topologies onto simpler interconnect topologies such as meshes or tori leads to difficult topology mapping and resource scheduling problems. A hybrid approach that employs optical circuit switches to reconfigure the interconnect topology to match application requirements can retain the advantages of a fully connected network using far fewer components. No timing models were used in this study whose focus was on the mapping of the inter-chip communication topologies rather than performance.



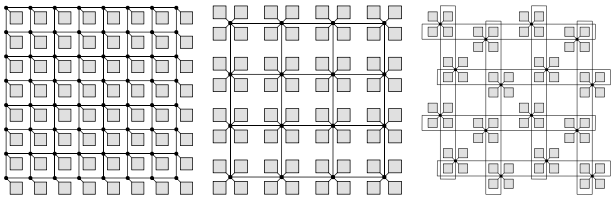
(a) Off state (b) On state (c) 4×4 switch

Figure 1. Photonic Switching Element. (a) Message propagate straight through. (b) Light is coupled into the perpendicular path. (c) A combination of eight ring resonators allows the construction of a 4×4 nonblocking optical switch.

3 Studied Network Architectures

This section describes the NoC architectures we examine which includes various networks for both conventional electronic networks and hybrid photonic-electronic networks.

3DI utilizing Thru-Silicon-Vias (TSVs) showcases inherently short interconnect paths with reduced resistance and capacitance, as well as lower power consumption. These characteristics enable the TSV's to enable the switching plane to be integrated on a separate plane of stacked silicon with very low power dissipation for the vias that connect between the planes. For the 32 nm technology node, the TSV is expected to scale to a 1.4 μm contact pitch, 0.7 μm diameter, almost 5×10^7 cm^{-2} maximum density, and 15 μm maximum layer thickness [10]. By stacking memory and interconnect resources on dedicated CMOS layers above the processors, it is possible to integrate larger memories and faster interconnects with future CMPs [17]. Silicon nanophotonic technology may alleviate the limitations of conventional electronic networks by using optics to deliver much higher bandwidth within the same power budget, however it has several inherent limitations, such as the inability to perform buffering and processing in the optical



(a) Mesh (b) Concentrated Mesh (c) Concentrated Torus

Figure 2. Mesh, concentrated mesh, and concentrated torus topology. The concentrated topologies require a larger-radix switch, but reduce the average hop count.

domain, which need to be circumvented in order to take the full advantage of this new technology.

Electrical NoC Architecture. We assume a CMP with 64 processors arranged in a 2D planar fashion. Although we do not simulate the processors themselves, we assume simple in-order cores with local store memories. The individual core size is $1.5mm \times 2.0mm$; the cores are located on the lowest layer of the 3DI CMOS die. Above the bottom layer are multiple layers devoted to the local store, allowing our cores sufficient capacity to feed computational units. Lastly, the top layer is where the global NoC is found. This consists of the electronic routers, and for the systems that include a photonic NoC, silicon nanophotonic components.

For our electrical network, we model the topologies shown in Figure 2. The mesh topology is the baseline for our comparisons against all of the other studied networks. In comparison to more exotic electronic networks, the mesh is simple to implement due to its use of relatively low radix switches in a regular 2D planar layout.

We also incorporate the concept of concentrating processing cores at a network node, originally explored in [1]. For example, a full mesh would include an access point for each node, creating an 8×8 mesh. By concentrating a set of four nodes together, the size of the mesh can be reduced to 4×4 thereby reducing the average hop count each message must incur but increasing the radix of each router to accommodate the four node connections. We explore the use of a concentrated mesh and concentrated torus, shown in Figure 2 (b) and (c). Note that unlike the concentrated networks in [1], the topologies we explore do not contain express channels between non-adjacent switches.

Photonic NoC Architectures. The photonic NoC is composed of two layers on the top plane of the 3DI structure, a photonic layer and an electronic control layer. The photonic layer provides a high bandwidth network for transmitting data and is constructed using silicon nanophotonic ring resonator structures that can be switched to control the propagation of optical signals (Figure 1). The electronic control layer is a secondary network used to transmit and act on control packets for the purpose of setting up and breaking down photonic links on the photonic layer. The control layer can also be provisioned as a low bandwidth network for transmitting small amounts of data.

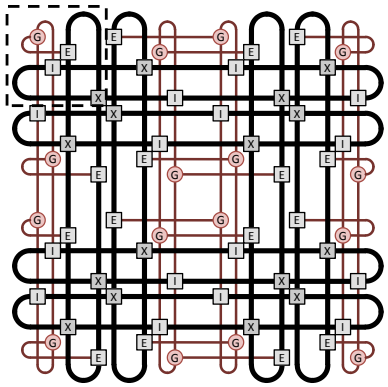


Figure 3. The photonic torus topology, studied in [15]. Switch blocks are abbreviated: X - 4×4 nonblocking, I - injection, E - ejection, G - gateway. Zoomed in view of the dotted box is shown in Figure 4.

Switching functionality on the photonic layer is derived from the use of ring resonator structures that act as PSEs, as in [15]. In Figure 1(a), the PSE is shown in the off-resonance state where messages propagate straight through the switch. Figure 1(b) shows the on-resonance state of the PSE, which bends the optical pathway implementing a turn. A control system is fabricated along with the switch to enable active switching of the device. The PSE models are implemented with the on-resonance state dormant, where no electrical current is applied, while the off-resonance state draws current to change the behavior of the device. By combining several PSEs together, functional network components such as the 4×4 nonblocking switch shown in Figure 1(c) can be created.

As described in [15], the main network structure of the topology is a folded torus shown as black lines in Figure 3. Included on the same topology is an additional set of waveguides and switches, shown as red lines, that are used to inject and eject optical messages into and from the network. Typically, this network provides a single access point for each processing node; however, we also include variations of this network with concentrated nodes, as previously described.

The transmission of data on the photonic network is enabled through the use of circuit switching, which requires the provisioning of an optical path before any data can be injected. The *path-setup phase* begins by sending a electronic setup control packet in the control layer, which travels through the network, establishing an optical path by

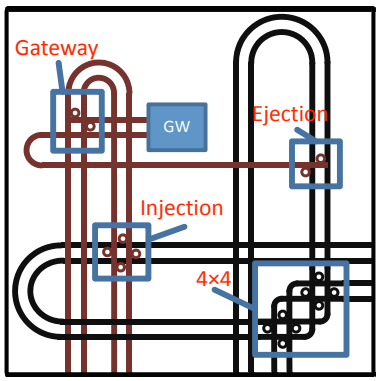


Figure 4. View of a single node in the photonic torus. The node(s) are connected to the gateway (GW) and the boxed areas represent switches used to control optical paths through the network.

configuring the appropriate PSEs. Once the setup packet reaches the destination node, the complete optical path has been allocated and an electronic acknowledgment is returned — allowing the source to begin data transmission upon receipt. The *breakdown phase* occurs upon complete transmission of data, where a breakdown control packet is sent along the network to release the optical path.

Figure 4 shows a detail view of the required photonic components needed to transmit and receive messages on the photonic NoC. The processing node (or nodes, for the concentrated configuration) injects messages electrically to the *gateway*, marked *GW*. Upon receiving an acknowledgement packet for a setup request, the gateway begins transmitting the message optically. The message first propagates through a *gateway switch*, which handles the routing of messages going to and from the gateway. Next, the message is directed towards the *injection switch* where it is switched into the torus network. The message then propagates through the torus (using dimension-ordered routing) until it reaches the correct turning point where it turns at a *4×4 nonblocking switch*. Once at the destination, the message exits the network via the *ejection switch*, and is directed to the gateway by the gateway switch where it is converted to an electronic signal and forwarded to the proper node.

Selective Transmission. Networks that transmit data exclusively on a photonic network ideally should favor large message sizes so that the path-setup overhead is sufficiently amortized over the transmission time of the entire message. Applications that send many small messages are subject to the full penalty of the path-setup overhead and will see substantially lower performance. In this study, we also include a *selective transmission* configuration of the photonic NoC that leverages the use of the electronic network as a low bandwidth data transmission medium. This configuration filters the packets using a size threshold, and transmits the data along the network that is most appropriate. A preliminary study using random traffic indicates a cross-over point of 256 bytes where transmitting smaller packets over the electronic control layer results in better performance and energy efficiency than using the photonic network alone.

4 Studied Benchmarks

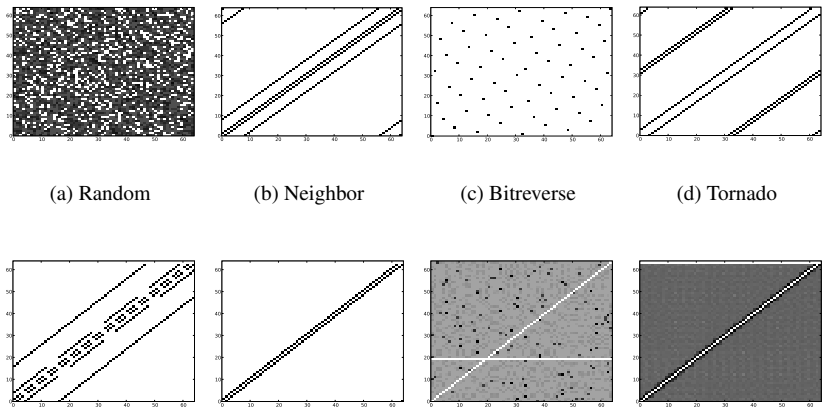
Our work extends related work by utilizing two sets of benchmarks: both standard synthetic traffic patterns and scientific application traces. Whereas the synthetic benchmarks help to identify the kinds of traffic best suited for each architecture, the application-based communication traces put real scientific workloads on the networks and test different mapping parameters. Figure 5 shows the spy plots of the eight benchmarks in this study. These plots illustrate the communication volume between each set of processors: a white square at the coordinate (p_i, p_j) in the plot represents no communication, while darker shades of gray represent increasing volumes of communication between two given processors. Details of the different benchmarks are given in Table 1.

Synthetic Benchmarks. We compare our NoC testbeds using four standard synthetic benchmarks from the literature [9], shown in the top of Figure 5. For each synthetic messaging pattern, two instances of the test are run: one with small messages and another with larger messages. Because of the restrictions of the hybrid interconnect studied, message transmissions are modeled as follows: each processor sends its messages as fast as possible, but blocks until receiving an acknowledgment from the destination processor before sending the next message.

In the *Random* test, each processor sends several messages to destinations chosen uniformly at random, independently from the previous destinations. *Neighbor* is a stan-

Table 1. Benchmark Statistics

Benchmark	Num Phases	Num Messages	Total Size (B)	Avg Msg Size (B)
Random-Small	1	6400	614400	96
Random-Large	1	6400	819200000	128000
Neighbor-Small	1	6400	614400	96
Neighbor-Large	1	6400	819200000	128000
Bitreverse-Small	1	6400	614400	96
Bitreverse-Large	1	6400	819200000	128000
Tornado-Small	1	6400	614400	96
Tornado-Large	1	6400	819200000	128000
Cactus	2	285	7296000	25600
GTC	2	63	8177148	129796
MADbench	195	15414	86516544	5613
PARATEC	34	126059	5457332	43.3



(a) Random (b) Neighbor (c) Bitreverse (d) Tornado

(e) Cactus (f) GTC (g) MADbench (h) PARATEC

Figure 5. Spyplots for the synthetic traces (top) and studied applications (bottom).

dard test where each processor sends messages to its neighboring processors in the physical two-dimensional topology of the NoC. The last two synthetic messaging patterns are designed to stress two-dimensional NoC topologies: the communication of the *Bitreverse* pattern requires each processor to send a message to its corresponding bitreversed address, involving traversals to far regions of the network. Lastly, *Tornado* is a pattern designed to stress 2D meshes by having each processor communicate to its neighbor's neighbors; the idea is to "shift" the communication of the Neighbor pattern in an adversarial way.

Each of the synthetic benchmark traces are generated from their descriptions in the literature using Python scripts.

Application-Based Benchmarks. A novel contribution of this research is the use of actual application communication information for the simulation of network performance. We developed a custom-designed profiling interface, used along with Linux's library preloading feature to overload the communication functions, thus keeping track of all function calls in an efficient, fixed-size array. At the end of application execution, we output our trace data to a separate file for each process, and the files are later combined. In order to accurately approximate communication behavior without including computation time, the trace tools order the communication into "phases" that are composed of sets of communications that must complete before further communication; essentially, we use the point-to-point synchronizations inherent in message passing to build an ordering of the communication.

We profile and study four different SPMD-style scientific applications, with traces obtained using a custom framework to measure interprocessor communication. The parallelization style of these applications is an ideal starting point for our study, because of their easily understandable synchronous communication model and their wide use in the scientific programming community.

The first evaluated application is *Cactus* [6], an astrophysics computational toolkit designed to solve coupled nonlinear hyperbolic and elliptic equations that arise from Einstein's Theory of General Relativity. Consisting of thousands of terms when fully expanded, these partial differential equations (PDEs) are solved using finite differences on a block domain-decomposed regular grid distributed over the processors. The Cactus communication characteristics reflect the requirements of a broad variety of PDE solvers on non-adaptive block-structured grids.

The Gyrokinetic Toroidal Code (*GTC*) is a 3D particle-in-cell (PIC) application developed to study turbulent transport in magnetic confinement fusion [13]. GTC solves the non-linear gyrophase-averaged Vlasov-Poisson equations in a geometry characteristic of toroidal fusion devices. By using the particle-in-cell method, the non-linear PDE describing particle motion becomes a simple set of ordinary differential equations (ODEs) that can be solved in the Lagrangian coordinates. GTC's Poisson solver is localized to individual processors, allowing the communication traces to only reflect the needs of the PIC core.

The PARAllel Total Energy Code [7] (*PARATEC*) is a materials science application that is widely used to study properties such as strength, cohesion, growth, and transport for materials like nanostructures, complex surfaces, and doped semiconductors using the Density Functional Theory (DFT) method. In solving the Kohn-Sham equations using a plane wave basis, part of the calculation is carried out in real space and the remainder in Fourier space using specialized parallel 3D FFTs. The all-to-all communication used to implement the 3D data transpose for the FFT is the most

demanding portion of PARATEC's communication characteristics.

Finally, we examine *MADbench* [4], a benchmark based on the MADspec cosmology code. MADspec calculates the maximum likelihood angular power spectrum of the cosmic microwave background (CMB). MADbench tests the overall performance of the subsystems of real parallel architectures by retaining the communication and computational complexity of MADspec and integrating a dataset generator that ensures realistic input data. Much of the computational load of this application is due to its use of dense linear algebra, which is reflective of the requirements of a broader array of dense linear algebra codes in scientific workloads.

Together, these four applications represent a broad subset of scientific codes with particular communication requirements both in terms of communication topology and volume of communication. For example, the nearest-neighbor Cactus communication represents components from a number of applications characterized by stencil-type behavior. Thus, the results of our study are applicable to a broad range of numerical computations.

5 Simulation Methodology

We have developed a comprehensive simulation framework capable of capturing key low-level physical details of both optical and electronic components, while maintaining cycle-accurate functional modeling using event-driven execution to achieve low-overhead simulation. The core framework is implemented in the OMNeT++ environment [19], which consists of around 25k lines of code, many of which are dedicated to specifying the detailed layout of photonic devices. Though OMNeT++ enables a modular construction and hierarchical instantiation of components, subtle differences in spatial positioning and orientation require some manual configuration of each network.

The electronic NoC, which is studied as a network for comparison, is functionally modeled cycle-accurately at 5 GHz. Electronic components, which pertain to both the electronic NoC and the electronic control plane of the photonic networks, are discussed below, followed by the photonic devices.

Processing Cores. Trace files captured from evaluated benchmarks (Section 4) are read into a processing core model that injects messages into the network. Messages are injected as quickly as possible for each messaging phase, once the core is finished with previous communication. This simulates the bulk-synchronous style of communication employed by the studied applications. Likewise, the destination processors take flits out of the network as soon as they arrive, under the assumption that the processor is not busy performing other computation or communication. This methodology is used to stress the network, illustrating the effects of having many messages in-flight. The trace files keep track of individual messaging phases in the application. Explicit small synchronization messages are sent to and from a master core, which enforces barriers between application phases.

In addition, communication elements are randomly assigned to cores in the network for the application data, to decrease the likelihood of a trace producing especially poor results by exploiting a single aspect of the network — a common artifact in real scientific computing. Each simulation is run fifty times with different mappings for each trace and topology, and the min, max, and average are subsequently collected. This randomization is not performed for the synthetic traces because they are intended to stress specific aspects of the physical NoC layout.

Routers. The router model implements XY dimension ordered routing with bubble flow control [16] for deadlock prevention and to avoid overrunning downstream buffers. Additionally, the routers are fully pipelined with four virtual channels and can issue two grant requests in a single cycle. For power dissipation modeling, the ORION electronic router model [21] is integrated into the simulator, which provides detailed technology-specific modeling of router components such as buffers, crossbars, and arbiters. The technology point is specified as 32 nm. Buffer sizes, shown in Table 2, are determined through preliminary experiments that identify optimal power-performance tradeoffs for each implementation to enable a fair comparison between electronic and photonic networks. In general, purely electronic networks have larger buffers and channel widths to increase their performance. This involves an important tradeoff with power consumption, making it necessary to gauge efficiency and not merely performance or power, which will be discussed further in the analysis of the results obtained. The concentrated networks also have larger buffers, presuming that this is appropriate given the smaller network size. Finally, the photonic networks using the *Selective* message filter have larger buffers to accommodate the electronic traffic that is allowed to travel on the interconnect.

Wires. Our detailed wire model is based on data collected for various wire lengths with different numbers of repeaters, running at 5 GHz with double pumping. This allows us to optimally buffer wires for power dissipation (around 50 fJ/bit/mm), which dictates the wire latency. Individual wire lengths are calculated using core size, router area (calculated by ORION), number of routers, and topology.

Photonic Devices. Modeling of optical components is built on a detailed physical layer library that we and others

Table 2. Electronic Router Parameters

Topology	Channel Width	Buffer Size (b)
Electronic Mesh	128	1024
Electronic Concentrated Mesh	128	2048
Electronic Concentrated Torus	128	2048
Photonic Torus	32	512
Selective Photonic Torus	64	1024
Photonic Concentrated Torus	32	1024
Selective Photonic Concentrated Torus	64	2048

have validated through the physical measurement of fabricated devices. The modeled components are primarily fabricated in silicon at the nano-scale, and include modulators, photodetectors, waveguides (straight, bending, crossing), filters, and PSEs consisting of ring resonators. These devices are characterized by attributes such as insertion loss, extinction ratio, delay, and power dissipation. Table 3 shows the optical parameters used [12, 20], excluding insertion loss and extinction ratio for brevity. Devices are sized appropriately and laid out into a network topology, which is controlled by the underlying electronic network.

A key parameter for the photonic devices, which greatly affects network performance, is the number of allowable wavelengths. This number is ultimately constrained by network size, since larger networks will exhibit a greater network level insertion loss [8]. The upper limit on available source power is the non-linear threshold of the ring resonators, while the lower limit in received power is dictated by the sensitivity of the photodetectors. An important advantage of our detailed simulator is the ability to perform this physical layer analysis, as shown in Figure 6, which determines the number of wavelengths available at different power budgets for a 64-core photonic torus. We found that 65 wavelengths can be used for the normal 8×8, and 150

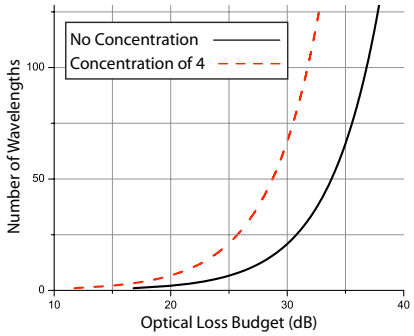


Figure 6. Insertion loss analysis of Photonic Torus topology.

Table 3. Optical Device Parameters

Sim Parameter	Value
Data rate (per wavelength)	10 Gb/sec
PSE dynamic energy	375 fJ*
PSE static (OFF) energy	400 uJ/sec†
Modulation switching energy	25 fJ/bit‡
Modulation static energy (ON)	30 μW§
Detector energy	50 fJ/bit¶
Wavelengths (8×8 network)	65
Wavelengths (4×4 conc. network)	128

for the 4×4 concentrated network for an optical power budget of 35 dB. We limit the max number of wavelengths to 128, considering space limitations on laser delivery to the modulators.

6 Results

We now evaluate the performance characteristics of the selected NoC implementations using the synthetic and application traces. The synthetic benchmarks provide a high-level picture of the interconnect's responsiveness to different commonly-observed communication patterns, while the application traces give insight to performance under realistic scientific loads.

The reported metrics are as follows: (1) performance is analyzed via the execution time of the benchmark or application, (2) energy cost by the total energy spent in execution, and (3) energy efficiency by the performance gained from each unit energy. Note that while typical network comparisons use message latency as a performance metric, such analysis would underscore the true performance of the system by only examining the transmission speed of single streams of data. Because the execution times and energies of the benchmarks varies broadly, we normalize the results to the electronic mesh performance. We choose an electronic mesh as the baseline because it represents the most straightforward engineering approach to interconnecting cores for emerging manycore processor designs.

Recall that the scientific application experiments are conducted using fifty random process placements to develop a statistical view of the networks responsiveness to varying communication mappings (see Section 5). Appli-

*Dynamic energy dissipation calculation based on carrier density, assuming 50-μm micro-ring diameter, 320-nm × 250-nm micro-ring waveguide cross-section, 75% waveguide volume exposure, 1-V forward bias.

†Based on switching energy, including photon lifetime for re-injection.

‡Same as *, for a 3μm ring modulator.

§Based on experimental measurements in [22]. Calculated for half a 10GHz clock cycle, with 50% probability of a 1-bit.

¶Conservative approximation assuming femto-farad class receiverless SiGe detector with $C < 1fF$.

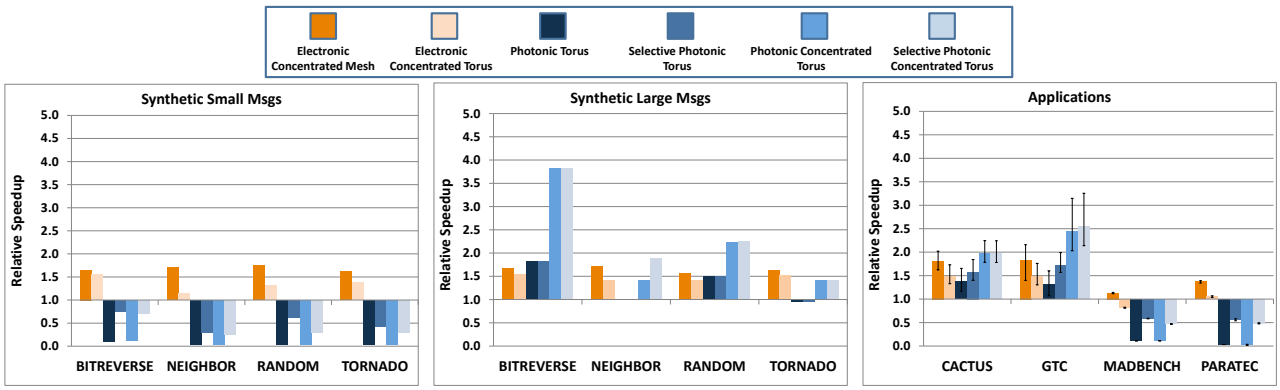


Figure 7. **Network speedup relative to the electronic mesh.**

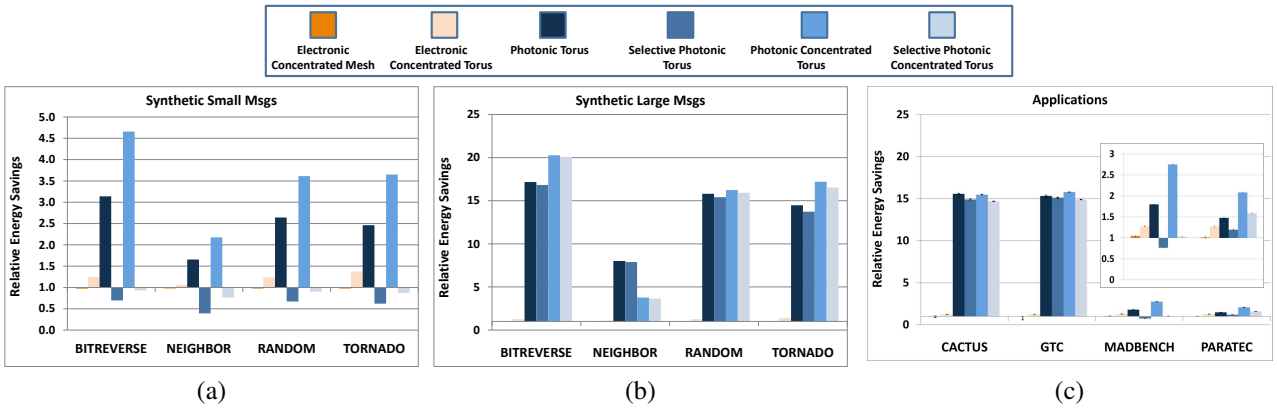


Figure 8. **Energy savings relative to electronic mesh. MADbench and PARATEC shown in inset for clarity in (c).**

cation results are therefore shown using the average performance, with error bars indicating min and max behavior.

Network Speedup. Figure 7 presents the application execution time speedup achieved by the examined NoC architectures relative to the execution time of the baseline electronic mesh. Values start at one, which indicates even performance with the baseline. For the synthetic tests with small messages, which are shown in Figure 7 (a), the photonic networks without selective transmission do not show improved performance, because the setup messages result in increased latency that is not sufficiently amortized by the high bandwidth end-to-end transmission of the photonic network. We see that selective transmission shows improvement, but does not gain in speedup over the electronic mesh due to the increased number of routers in the hybrid network used for injection and ejection (see Figure 3). The synthetic tests with large messages, which are displayed in Figure 7 (b), show a significant improvement for the hybrid photonic networks, compared to what is observed for the experiments conducted on small messages. This illustrates the benefit of amortizing the setup overhead for purely circuit-switched photonic networks. Additionally, it is interesting to note the improvement for the Bitreverse benchmark, which exhibits significantly longer communication patterns, in that circuit-switching directly improves the performance by mitigating contention on a one-time basis. Recall that the effective bandwidth of the photonic network only matches that of the electronic ones when the photonic network is concentrated ($128\lambda \times 10$Gbps vs. 128 channel width $\times$ 5GHz double pumped), which is why they perform significantly better than their full-network counterparts.

Figure 7 (c) shows the relative speedup of the real application traces. The concentrated photonic networks clearly outperform the other interconnect configurations for both Cactus and GTC, similar to the synthetic large-message traces. The photonic networks do not perform as well for the MADBench and PARATEC applications primarily because those benchmarks exhibit all-to-one and broadcast communication patterns, which are expected to behave poorly in circuit-switched networks. For these types of applications, wavelength-routed inter-core networks would likely be more appropriate, and future work investigating the use of both circuit-switched and wavelength-routed photonics is under way. In addition, these two benchmarks use significantly smaller message sizes (see Table 1). The se-

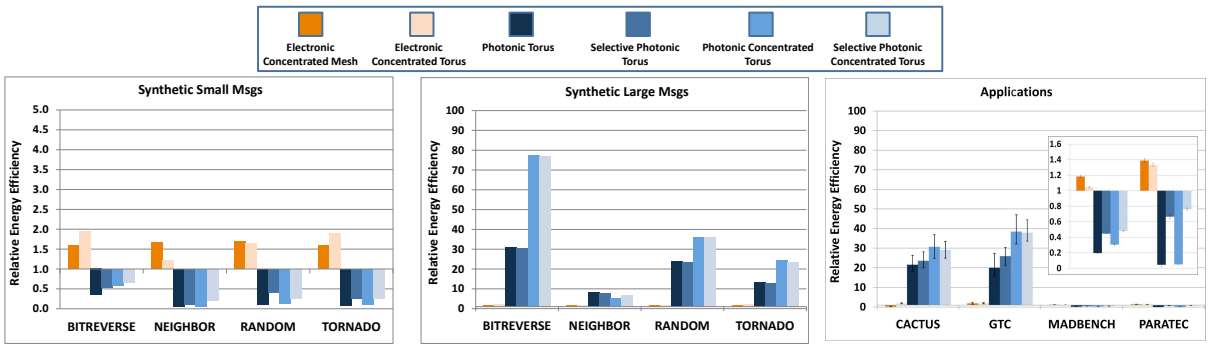


Figure 9. Energy efficiency (network performance per unit energy) relative to the electronic mesh. MADbench and PARATEC shown in inset for clarity in (c).

lective networks narrow the performance difference somewhat, but still do not achieve the nominal performance of the electronic mesh network, similar to the synthetic traces using small messages.

Energy Consumption. Figure 8 presents the results of the metric of total energy consumption; the plot shows the inverse of consumption (i.e. the energy savings), again relative to the electronic mesh baseline. The photonic networks are clear winners for most experiments — particularly the large-message synthetics as well as Cactus and GTC applications — showing over 10× improvement due to the decoupling of distance, bandwidth, and power during optical transmission. Since the circuit-switched photonic network does not consume power per-hop, the energy usage is much lower than the packet-switched electrical networks, which require energy consumption in order to make routing decisions at each hop. This point is particularly illustrated again in the Bitreverse benchmark. Because photonics is completely decoupled from distance travelled with repect to energy spent during transmission, it will provide higher benefits when communication pairs are further apart.

Performance for Energy Spent. Figure 9 shows the final metric: performance gained for every unit of energy spent, which is effectively a measure of a network's efficiency. This metric is calculated by multiplying the network execution time by the energy spent (plotted as the inverse so that values greater than 1 indicate a better performance per energy). The numbers are shown relative to the electronic mesh.

The benchmarks with small messages perform poorly on photonic networks, as seen in Figure 9 (a). Although network speedup is reasonable for some photonic networks in Figure 7, and energy gains are achieved for some photonic networks in Figure 8, the overall network performance is not improved over the electronic mesh when message sizes are small.

However, as shown in Figures 9 (b) and (c), the photonic networks' energy efficiency improvement over the electronic mesh for traces with large message sizes is amplified by the gains in both speedup and energy, resulting in improvements of over 20×. This benefit is realized over a variety of communication patterns, including two of the real applications, which demonstrates the possible appeal of on-chip photonics for many classes of applications.

7 Conclusions and Future Work

This work compares the performance and energy characteristics of electronic and photonic NoCs using a suite of synthetic communication benchmarks as well as traces from SPMD-style scientific applications on a detailed simulation framework. We show that a hybrid NoC has the potential to outperform electrical NoCs in terms of performance, while mitigating the power/energy issues that plague electronic NoCs when the communications are sufficiently large to amortize the increased message latency. For messaging patterns with small messages and high connectivity, the current photonic network design does not perform as well as an electronic mesh, although parameter searches may mitigate this by sizing queues and message size cutoffs to enable better performance in the selective approach.

The comprehensive and detailed level of simulation as well as the range of applications and topologies investigated achieves interesting results that are not possible using a higher-level analysis. These observations will be important in guiding future CMP engineers who seek to design an interconnect architecture that does not become the bottleneck for performance or energy. As future architectures scale to even higher concurrencies, the power requirements and performance benefits of photonic interconnects will become increasingly attractive.

Although these results have addressed some questions about how different applications would behave on different NoCs, it also raises a number of concerns that will lead to

important future studies. This work focuses completely on the interconnection network and does not account for data transfer onto the chip from DRAM, nor does it account for computing performance. Furthermore, it is not clear how the performance and energy consumption of the networks fit into overall system performance and energy, and how communication can be overlapped with computation more efficiently. These experiments are currently being pursued for future work.

Alternative topologies for both electronic and photonic networks must also be explored. Photonic network architectures that exhibit less blocking under heavy loads have been proposed in related work, and will be examined in detailed future studies. Many methods of improving electronic interconnect performance are also emerging that may substantially change the comparison between photonic and electronic NoCs.

A key contribution of our work was the focus on SPMD style applications found in the scientific community. Although many elements of these algorithms are finding their way into consumer applications such as realistic physics for games, and image processing kernels, future studies will also explore applications with more asynchronous communication models. We plan to make a deeper examination of the differences between message passing and shared memory applications and how they interact with both photonic and electronic networks characteristics. All of these refinements will be subjects for future work, using the foundation presented in this paper.

8 Acknowledgements

This research is partially supported by DARPA MTO office under grant ARL W911NF-08-1-0127 and the National Science Foundation (Award #: 0811012).

References

[1] J. Balfour and W. Dally. Design tradeoffs for tiled CMP on-chip networks. In *International Conference on Supercomputing*, 2006.

[2] C. Batten et al. Building manycore processor-to-DRAM networks with monolithic silicon photonics. In *Proceedings of 16th IEEE Symposium on High Performance Interconnects*, Aug 2008.

[3] K. Bernstein et al. Interconnects in the third dimension: Design challenges for 3D ICs. In *Design Automation Conference*, 2007.

[4] J. Borrill et al. Integrated performance monitoring of a cosmology application on leading HEC platforms. In *International Conference on Parallel Processing (ICPP)*, 2005.

[5] M. Briere et al. Heterogeneous modeling of an optical network-on-chip with SystemC. In *16th IEEE International Workshop on Rapid System Prototyping*, 2005.

[6] Cactus, 2004. http://www.cactuscode.org.

[7] A. Canning, L. Wang, A. Williamson, and A. Zunger. Parallel empirical pseudopotential electronic structure calculations for million atom systems. *Journal of Computational Physics*, 160:29–41, 2000.

[8] J. Chan, A. Biberman, B. G. Lee, and K. Bergman. Insertion loss analysis in a photonic interconnection network for on-chip and off-chip communications. In *IEEE Lasers and Electro-Optics Society (LEOS)*, Nov. 2008.

[9] W. Dally and B. Towles. *Principles and Practices of Interconnection Networks*. Morgan Kaufmann Publishers, 2004.

[10] The international technology roadmap for semiconductors (ITRS). http://www.itrs.net.

[11] S. Kamil, A. Pinar, D. Gunter, M. Lijewski, L. Oliker, and J. Shalf. Reconfigurable hybrid interconnection for static and dynamic applications. In *ACM International Conference on Computing Frontiers*, 2007.

[12] B. G. Lee et al. High-speed 2×2 switch for multi-wavelength message routing in on-chip silicon photonic networks. In *European Conference on Optical Communication (ECOC)*, Sept. 2008.

[13] Z. Lin, S. Ethier, T. Hahm, and W. Tang. Size scaling of turbulent transport in magnetically confined plasmas. *Physical Review Letters*, 88, 2002.

[14] I. O'Connor et al. Towards reconfigurable optical networks on chip. In *Reconfigurable Communication-centric Systems-on-Chip workshop*, June 2005.

[15] M. Petracca, B. G. Lee, K. Bergman, and L. Carloni. Design exploration of optical interconnection networks for chip multiprocessors. In *16th IEEE Symposium on High Performance Interconnects*, Aug 2008.

[16] V. Puente, R. Beivide, J. A. Gregorio, J. M. Prellezo, J. Duato, and C. Izu. Adaptive bubble router: a design to improve performance in torus networks. In *Proc. Of International Conf. On Parallel Processing*, pages 58–67, 1999.

[17] A. Shacham, K. Bergman, and L. P. Carloni. Photonic networks-on-chip for future generations of chip multiprocessors. *IEEE Transactions on Computers*, 57(9):1246–1260, 2008.

[18] D. Vantrease et al. Corona: System implications of emerging nanophotonic technology. In *Proceedings of 35th International Symposium on Computer Architecture*, Aug 2008.

[19] A. Varga. OMNeT++ discrete event simulation system. http://www.omnetpp.org.

[20] Y. Vlasov, W. M. J. Green, and F. Xia. High-throughput silicon nanophotonic wavelength-insensitive switch for on-chip optical networks. *Nature Photonics*, 2:242–246, April 2008.

[21] H. Wang et al. ORION: A power-performance simulator for interconnection networks. In *35th International Symposium on Microarchitecture*, 2002.

[22] M. R. Watts. Ultralow power silicon microdisk modulators and switches. In *5th Annual Conference on Group IV Photonics*, 2008.