F^{2}BFLY: An On-Chip Free-Space Optical Network with Wavelength-Switching

Jin Ouyang[†], Chuan Yang[§], Dimin Niu[†], Yuan Xie[†], Zhiwen Liu[§]
[†]Dept. of Computer Science and Engineering*
[§]Dept. of Electrical Engineering
Pennsylvania State University, University Park, PA 16802
jouyang@cse.psu.edu, czy111@psu.edu

ABSTRACT

The increasing number of cores in contemporary and future many-core processors will continue to demand high throughput, scalable, and energy efficient on-chip interconnection networks. To overcome the intrinsic inefficiency of electrical interconnects, researchers have leveraged recent developments in chip photonics to design novel optical network-on-chip (NoC). However, existing optical NoCs are mostly based on passively switched, channel-guided optical interconnect in which large amount of power is wasted in heating the micro-rings and maintaining the optical signal integrity.

In this paper we present an optical NoC based on *free-space* optical interconnect in which optical signals emitted from the transmitter is propagated in the free space in the package. With lower attenuation and no coupling effects, free-space optical interconnects have less overheads to maintain the signal integrity, and no energy waste for heating micro-rings. In addition, we propose a novel cost-effective *wavelength-switching* method where a refractive grating layer directs optical signals in different wavelengths to different photodetectors without collision. Based on the above interconnect and switching technologies, we propose **free flattened butterfly** (F^{2}BFLY) NoC which features both high-radix network and dense free-space optical interconnects to improve the performance while reducing the power. Our experiment results, comparing F^{2}BFLY with state-of-the-art electrical and optical on-chip networks, show that it is a highly competitive interconnect substrate for many-core architectures.

Categories and Subject Descriptors

B.4.3 [**Hardware**]: Interconnections (Subsystems)

General Terms

Design, Experimentation

Keywords

free-space optical interconnects, network-on-chip

*J. Ouyang, D. Niu, and Y. Xie are supported in part by NSF CCF-0903432, CNS-0905365, SRC grant, and DoE's ASCR program under under award number DE-SC0005026.

ICS'11, May 31–June 4, 2011, Tucson, Arizona, USA.

1. INTRODUCTION

In an era when shrinking device size and exploiting instruction-level parallelism (ILP) gain diminishing returns in frequency and application throughput, many-core architectures which better leverage high device density and task-level parallelism (TLP) is proposed as a more efficient solution to continuous performance improvement. The increasing number of cores in a chip, however, places rapidly increasing communication burden on the on-chip interconnect substrate. Therefore, high throughput, scalable, and energy-efficient on-chip interconnection networks become the key enabler of future large scale many-core processors. Towards this objective, researchers have been focused on either designing novel router architectures and network topologies, or leveraging emerging on-chip interconnect technologies which are intrinsically more efficient than conventional electrical interconnects.

1.1 Electrical NoCs

A major branch of contemporary NoC researches inherit ideas from the long-established computer and multiprocessor networks [1]. However, the on-chip environment adds extra constraints and opportunities, which leads to various specially optimized architectures. Driven by the requirements of low latency and low complexity, researchers have been working on reducing the router pipeline stages and efficient switching mechanisms [2, 3, 4]. An interesting recent trend is to leverage the ample on-chip wiring resources to build high-radix, richly interconnected NoCs that features both high bi-section bandwidth and low network diameter [5, 6, 7, 8]. Grot *et al.* [7] generalize these high-radix NoCs as *express cube topologies* and perform a comprehensive study of the design space of express cubes. Their results show that by wisely utilizing the abundant wiring resources, express cubes can produce significantly higher performance and consume lower power than conventional topologies.

However, future performance elevation with electrical NoC is impeded by the slow improvement in the on-chip interconnect technology. The wiring density is not increasing as fast as device density, since to maintain acceptable signal integrity, the feature size of wires cannot scale at the

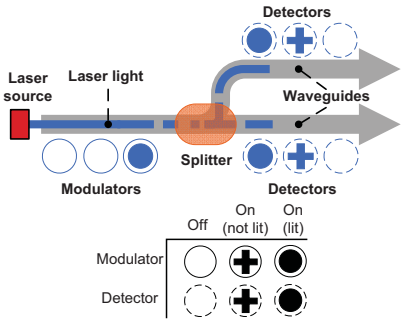


Figure 1: Channel-guided optical interconnects: a continuous laser beam is provided by the external laser source, propagated in on-chip waveguides, and modulated and detected by micro-rings.

same pace as devices. In addition, the increase of metal levels has almost stopped, and it is projected that from now to the 11nm node by 2022, only 4 more metal levels will be added (from 11 to 15 levels) [9]. Furthermore, the power spent by on-chip interconnects starts to dominate the overall chip power [10, 11]. The express cube topologies will suffer even more from above issues as the wiring intensity of these topologies increases with the network size ($O(N)$ for flattened butterfly [5] and $O(\sqrt{N})$ for MECS [6, 7], where N is the number of nodes in the network). This sets researchers on the quest of more scalable and power-efficient alternatives for on-chip interconnects.

1.2 Optical NoCs

Recent developments in nanophotonics have motivated researchers to design NoCs based on optical interconnects, which is intrinsically faster and less power consuming than electrical interconnects. The proposed optical NoCs [12, 13, 14, 15, 16, 17, 18] are mostly based on *channel-guided optical interconnects*, where optical signals are confined and propagated in on-chip waveguides.

1.2.1 Channel-Guided Optical Interconnects

Figure 1 provides a conceptual view of channel-guided optical interconnects and illustrates the necessary components. On-chip waveguides (fabricated with poly-Si, crystalline Si [14, 15], or polymer [19]) serve as the physical channel carrying optical signals. A continuous laser beam is provided by an external laser source and modulated into on-off digital signals. The modulated light continues to travel downstream, possibly duplicated by a splitter, and finally detected and removed from the waveguides. Modulation and detection are both performed by micro-rings placed beside the waveguides, which can be electrically tuned into and out of resonance with a certain wavelength. In the resonant state, the coupling between the micro-rings and the waveguides becomes so strong that the light is dropped by the micro-rings. Therefore, electrically tuning the micro-rings achieves on-off modulation, and dropping the optical signals to a photodetector completes detection. To further increase the bandwidth, dense-wavelength-division-multiplexing (DWDM) is used to enable the waveguide to carry multiple wavelengths simultaneously. The micro-ring can selectively modulate or drop a given wavelength by adjusting its resonance frequency.

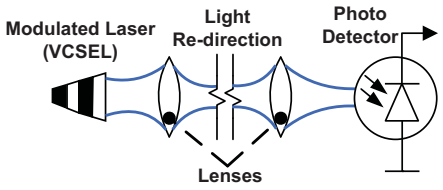


Figure 2: Free-space optical interconnects: the actively modulated VCSEL emits encoded optical signals, which are then collimated and redirected by a series of optical devices, and finally detected by the photodetector.

While channel-guided optical interconnects with DWDM are shown to be more energy-efficient than electrical interconnects, it has significant energy wastes that offset the energy-efficiency of optical interconnects. First, the external laser source needs to continuously produce lights used for modulation. To make things worse, existing optical NoCs typically have a large number of long waveguides due to the lack of optical buffering. This requires the external laser source to provide high laser power on the order of tens of watts. Second, to achieve high bandwidth thousands of micro-rings are used as modulators and detectors. Thermal trimming is needed to ensure all the micro-rings to work with the right resonance frequencies. Besides technical difficulties, the total power needed for trimming is so high to be comparable with the laser source power. Analysis shows that the laser source power and the trimming power accounts for 75% of total power in channel-guided optical networks [20]. Note that the above two power components are static powers and pure overheads.

1.2.2 Free-Space Optical Interconnects

Most recently, free-space optical interconnects (FSOI) gains interests as it overcomes the inefficiencies of channel-guided optical interconnects. Figure 2 shows a conceptual view of FSOI. Compared with channel-guided interconnects, the major differences include:

- **Laser source and modulation:** Light is emitted from on-chip integrated vertical-cavity surface-emitting lasers (VCSELs) [21, 22, 23, 24]. By switching on and off the VCSEL, active modulation is achieved with the laser source.
- **Light propagation:** Modulated lightwave propagates in the free-space in the package. In order to direct lightwave onto intended photodetectors, optical elements are inserted to achieve multiple functionalities of collimation, diffraction, and collection of light signal. We will discuss more about how to control the light's direction below and in the following section.
- **Detection:** Instead of using micro-rings to couple light out of waveguides for detection, photo diodes are employed. Light beams are projected onto corresponding photo diodes under the control of optical elements.

From the above discussion, we can see that free-space optical interconnects eliminate the needs of external laser source and micro-rings, which are the sources of significant energy wastes. In addition, optical signals propagating in the free space experience no coupling and lower attenuation, which further improves the efficiency of optical links.

However, new challenges are also introduced: directing the emitted laser beam to the intended photodetector. As exist-

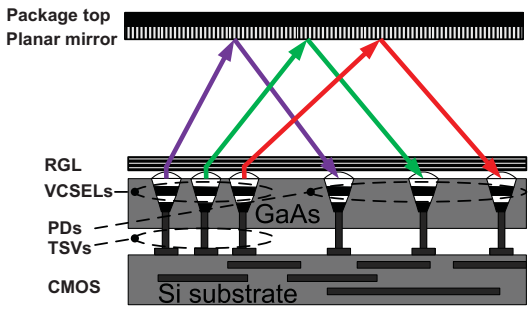


Figure 3: The optical system with 3D stacked RGL, optical substrate, and CMOS substrate.

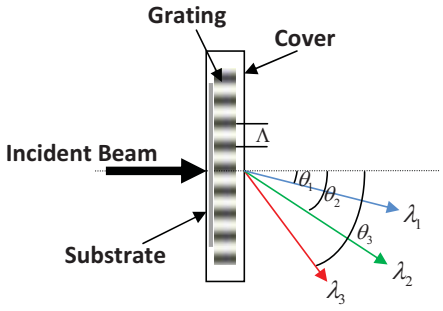


Figure 4: The volume phase holographic grating VPHG used as RGL.

ing VCSELs and photo diodes are all surface devices, they cannot be freely aimed at each other. Therefore, optical elements are placed in the optical path to control the direction of beam propagation. Xue *et al.* [25] propose to use angled micro-mirrors in the package to re-direct the beams. However, this method suffers from technological difficulties and complexities. Fabricating micro-mirrors requires unconventional processes that are incompatible with planar CMOS processes. To make things worse, aligning the micro-mirrors requires precise control of 5 free variables in the 3D space (x, y, z, inclination angle θ, and azimuth angle ϕ), whose feasibility remains a question. In addition, Xue *et al.* propose an all-to-all network where a laser beam will be reflected by multiple mirrors before reaching the destination. Planning the optical paths for all links in the complex all-to-all network is an intimidating task and adds more difficulties to alignment. Last but not the least, this also introduces problems to designing network architectures, which will be detailed in Section 3.

In contrast to previous approaches, we propose to achieve beam redirection with a single **Refractive Grating Layer** (RGL). A RGL is a planar thin-film diffractive optical element. One interesting feature of RGL is that the refractive angle (the extent to which the light direction is changed) depends on the wavelength of the incident light. That is, controlling the optical path is as simple as selecting the wavelength.

Figure 3 provides the high level view of our optical system. We leverage 3D integration to stack the optical substrate (GaAs) on CMOS substrate (Si). On the surface of the optical substrate, a planar RGL is in turn stacked. Laser beams generated by VCSELs are vertically incident onto the RGL, by which the laser beam get collimated and skewed. A planar mirror is deposited on the ceiling of the package, which bounces the beams back towards the corresponding photodetectors on the optical substrate. The advantages of this approach are: *(a)* the optical elements for beam redirection (planar mirror and RGL) can be fabricated with mature processes and *(b)* we only need to control 3 free variables (wavelength λ and 2D locations of VCSELs and detectors) to ensure that the laser beams are projected onto intended detectors. The relative positions of VCSELs and detectors can be precisely defined when fabricating the optical substrate; wavelengths λ is well controlled in contemporary multi-wavelength VCSELs [21, 22, 23, 24]. With the absence of alignment issues, the proposed optical system is more feasible to realize. Detailed analysis of the system is provided in Section 2.

Free Flattened Butterfly (F^2BFLY): We leverage the above proposed free-space optical interconnects to design an optical NoC called free flattened butterfly. Based on the special features of underlying interconnect technology, F^2BFLY features high-radix routers, rich connectivity, and low network diameter by fully exploiting the low latency and low power benefits of free-space optical interconnects. As we will show by experiment results, F^2BFLY outperforms electrical NoCs with both conventional mesh topology and high-radix express cube topologies. In addition, it is also more energy-efficient than state-of-the-art channel-guided optical NoCs. In designing the NoC architecture, we also propose optimizations to reduce the complexities of high-radix routers, resulting in a simple and low cost router suitable for on-chip implementation.

In the rest of the paper, we first provide in-depth description of the proposed free-space optical interconnects based on wavelength-switching (Section 2). We then present our architectural design in Section 3. To perform a comprehensive evaluation, we also investigate state-of-the-art electrical and optical NoCs together with our proposal, and introduce the simulation environment in Section 4. The simulation results are presented and discussed in Section 5. Finally, related work is reviewed and conclusion is drawn in Section 6 and 7.

2. DESIGNING FREE-SPACE OPTICAL INTERCONNECTS

In this section, we provide specific design considerations for the free-space optical interconnects illustrated in Figure 3, with a focus on enabling wavelength-switching with a high efficiency RGL:

• **RGL Design:** RGL is a thin-film component with periodic phase structure, which diffracts different wavelengths of light from a common input direction into different angular output directions (Figure 4). In this paper, we use the volume phase holographic grating (VPHG) as RGL due to its advantages in diffraction efficiency and system packaging [26, 27]. Figure 4 shows the structure of a VPHG. The core of a VPHG is a transmissive material whose refractive index is a periodic function of location, with a period of Λ. When a light beam passes through this structure, it is diffracted (or redirected) and the angle of the output beam θ depends on the wavelength λ and period Λ, as given by $\sin(\theta) = \lambda/\Lambda$. For example, in Figure 4 the incident light beam contains 3 different wavelengths which are redirected to different directions. In addition, the core of a VPHG is

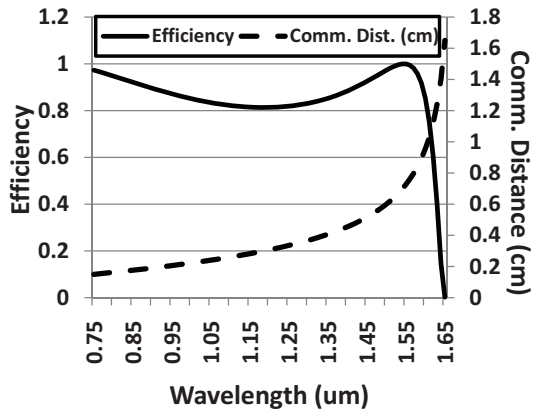


Figure 5: Diffraction efficiency (solid line) and communication distance (dash line) with respect to wavelength.

laminated by plates of glass or fused silica, which increase its durability for handling and system packaging.

The diffraction efficiency of the VPHG is high, reported to be better than 80% or even close to 100% [26, 27]. Based on the two-wave coupling theory, the diffraction efficiency depends on the grating thickness, the grating period, and the wavelength. Targeting at optical interconnect lengths in the range of 1.5mm to 12mm, we set the grating period to 1600nm (660 l/mm) and can achieve an efficiency greater than 75%. The diffraction efficiency and communication distance with respect to wavelength is plotted in Figure 5.

The index pattern of VPHG is generated by laser exposure, in a similar way to patterning silicon substrate. However, a single exposure run can only define one type of pattern, ruling out the possibility of defining arbitrary patterns on a single VPHG. Therefore in this work we only assume two types of patterns on the VPHG used: an x-pattern enabling x-dimension links and a y-pattern enabling y-dimension links (Figure 6). This restriction presents both challenges and opportunities for designing an free-space optical network, as we will discuss in detail in Section 3.

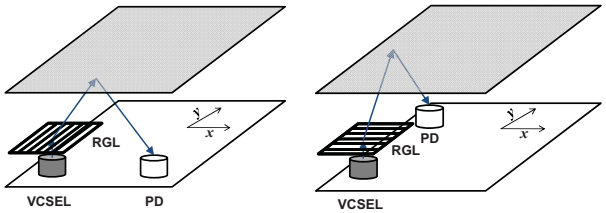
• **VCSELs and Photodetectors:** Contemporary integrated VCSELs feature small footprints on the order of tens of microns, high modulation speeds up to 40Gbps [28], and the ability to produce a single-wavelength laser beam from a wide range of wavelengths [21, 22, 23, 24]. The resonant cavity photodiodes incorporates the resonant cavity similar to VCSELs to amplify the received signal, and offers high sensitivity and bandwidth [29]. For the free-space optical links in this paper, we assume the high-speed VSCEL and photodetector design in [25]. The key parameters of the overall optical system is listed in Table 1.

3. FREE FLATTENED BUTTERFLY NOC

Leveraging the FSOI based on wavelength-switching in the previous section, we design F^{2}BFLY with realistic hardware implementation in mind and its architecture is detailed in this section.

3.1 Topology

While an all-to-all network seems attractive to efficiently use optical interconnects, practical issues mentioned in the



(a) Laser beam directed to the x-dimension (b) Laser beam directed to the y-dimension

Figure 6: The two RGL patterns directing laser beams to two orthogonal directions.

Table 1: Optical System Design

Package Height	1.5mm
Communication Distances	
Max.	12mm
Min.	1.5mm
RGL	
Wavelength	.75–1.6um
Grating density	660 lines/mm
Grating efficiency	$\geq$ 75%
VCSEL [25]	
Frequency	40GHz
Power	6.3mW
Area	20×20um^2
Photodetector [25]	
Frequency	40GHz
Power	4.2mW
Area	20×20um^2

previous section prevent implementing FSOI links with arbitration directions. In addition, the hardware complexity per node of an all-to-all network grows in the order of $O(N)$, which will become difficult to accommodate on-chip when the network scale goes up.

Instead, we propose a topology that both exploits optical efficiency and has good scalability. First, the FSOI links in our system can have only two possible directions, chosen from the cardinal directions (x and y directions). This eliminates the complexities associated with building links with arbitrary directions. In addition, coupled with a dimensionally decomposed router architecture (see Section 3.2), RGL fabrication and alignment difficulties are further reduced.

Using only the two cardinal directions, a node in the network can reach any node in the same row or column with wavelength-switching as discussed in the previous section.

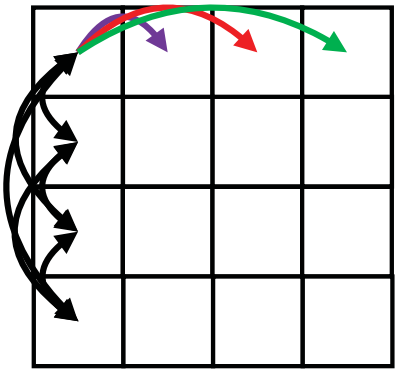


Figure 7: 1-d FSOI links and crossbar formed by wavelength-switching.

Figure 7 shows an example of a 16-node network, where 3 FSOI links using 3 different wavelengths from the node at the upper-left corner to all other nodes in the first row are shown. With FSOI links connecting any pair of nodes in the same dimension, a 1-dimensional crossbar is formed by the FSOI links for each row/column. For example, Figure 7 shows a 1-dimensional crossbar formed in the first column. Overall, the resulted network topology is equivalent to a 2-level flattened butterfly (FBFLY) [5], where the crossbars along the rows form the first level and those along the columns form the second level. With this topology, packets sent between nodes in the same row or column can reach the destination in one hop. For a packet sent to a node in both a different row and a different column, it is first routed in the first level (row dimension) and then in the second level (column dimension) to reach the destination. Therefore, any packet can reach the destination in at most 2 hops (the network diameter is 2 hops). Compared to all-to-all optical networks F^{2}BFLY introduces only one more hop and one more pair of O/E and E/O conversions, while both the implementation difficulty and hardware complexity per node (now $O(\sqrt{N})$) become significantly lower. In addition, it enables an efficient router architecture suitable for on-chip implementation, which will be detailed in the following subsection.

3.2 Router Architecture

Unlike 2D mesh whose routers have a constant number of output ports, F^{2}BFLY requires the output port counts of routers to increase in the order of $O(\sqrt{N})$. For example, a 16-node and a 64-node F^{2}BFLY requires routers with 7 and 15 output ports, respectively, including the ejection ports. Simply adopting the generic wormhole-switching router architecture [2] for such high-radix routers is inefficient, since the the hardware complexity of generic wormhole-switching routers grows quadratically with the radix. To design an efficient, high-radix router architecture for F^{2}BFLY, we propose to use *dimension-decomposition* to reduce the router complexity. In addition, we use only one virtual channel per input port and source routing to further simplify the router pipeline.

• **Dimension-Decomposition:** The idea of dimension-decomposition is to divide the router into a row module and a column module [30, 31], as shown in Figure 8a. As a result, the original 5×5 crossbar of the router is decomposed into 2 2×2 crossbars inside the row and column modules respectively. The incoming flits are guided to the row/column modules in order to traverse in the row/column dimension. Since the complexity of the crossbar grows quadratically with its port counts, the decomposed crossbars have lower hardware cost and power consumption than the original crossbar. Using the same approach, we can reduce the crossbar complexity of routers in F^{2}BFLY. For example, the 7×7 crossbar for a 16-node F^{2}BFLY can be decomposed into 2 4×4 crossbars whose overall complexity is lower.

There are further opportunities for optimization if we assume XY-routing is used. First, note that with XY-routing a flit can only turn from the row dimension to the column dimension but not reversely. Second, since in F^{2}BFLY a flit only takes at most 1 hop in each dimension, an incoming flit that is currently traversing in the row/column dimension will *not* continue to traverse in the same dimension. That is, an incoming flit using the row dimension will either turn to the column dimension or go to the ejection port; an incoming flit using the column dimension will definitely go to the ejection port (it cannot turn to the row dimension with XY-routing). Based on these observations, the complexities of switching fabrics in each dimension module can be further reduced.

The resulted overall router architecture for 16-node F^{2}BFLY is shown in Figure 8b. For clarity, input buffers are omitted in the figure. The router is decomposed into a row and a column module. Each module has a transceiver block (Tx-Rx Block in the figure) and switching fabrics (crossbar or multiplexers). The transceiver block consists of VCSEL and photodetector arrays, which forms the input and output FSOI links for the corresponding dimension. For a 16-node F^{2}BFLY, each transceiver block has 3 input ports and 3 output ports. The switching fabric directs an incoming flit to one of the output ports of the transceiver block or to the ejection port. The switching fabric of the column module is a 4×4 crossbar, whose inputs are the injection port and the 3 input ports from the transceiver block in the row module, and whose outputs are the ejection port and the 3 output ports to the transceiver block in the column module. The switching fabric of the row module contains only one 3-to-1 multiplexer and one 1-to-3 de-multiplexer. The inputs to the multiplexer are the input ports from the transceiver block in the column module and the output is the the ejection port. The de-multiplexer, on the other hand, takes the injection port as the input and the outport ports to the transceiver block in the row module as the outputs. Since a flit can only turn from the row module to the column module and cannot continue traversing a same dimension, the optimized router architecture can maintain the connectivity of F^{2}BFLY while using low-complexity switching fabrics instead of a 7×7 crossbar.

• **Router Pipeline and Source Routing:** As network deadlock is already avoided with XY-routing, our F^{2}BFLY router uses only 1 virtual channel which removes virtual channel allocation (VA) stage from the pipeline. In addition, the fact that the network diameter is 2 hops motivates us to use source routing to eliminate the route computation (RC) stage. In source routing, the source of the packet computes the output port to use at each hop during the injection process and encodes this information in the packet header. In F^{2}BFLY, the packet header only needs to record 2 output ports along the path, which introduces very low overhead. For example, the first half of Figure 9 compares the routing information contained in the header for per hop routing and source routing respectively, in a 16-node F^{2}BFLY. For source routing, the first bit (the FD bit) of routing information indicates whether to traverse the row dimension or the column dimension first. The two pairs of bits of the following 4 bits (OP bits) record the output ports to use when traversing the row dimension and the column dimension respectively. If the FD bit indicates that the first dimension to traverse is column dimension, the output port information for the row dimension is ignored. As we can see, the source routing information is only 1 bit longer than the per hop routing information, adding negligible overheads to the header. Finally, the second half of Figure 9 shows the resulted router pipeline where only the switch arbitration (SA) and switch traversal (ST) stages are left.

• **Flow Control:** We use credit-based flow control with F^{2}BFLY to ensure no packet is dropped when being trans-

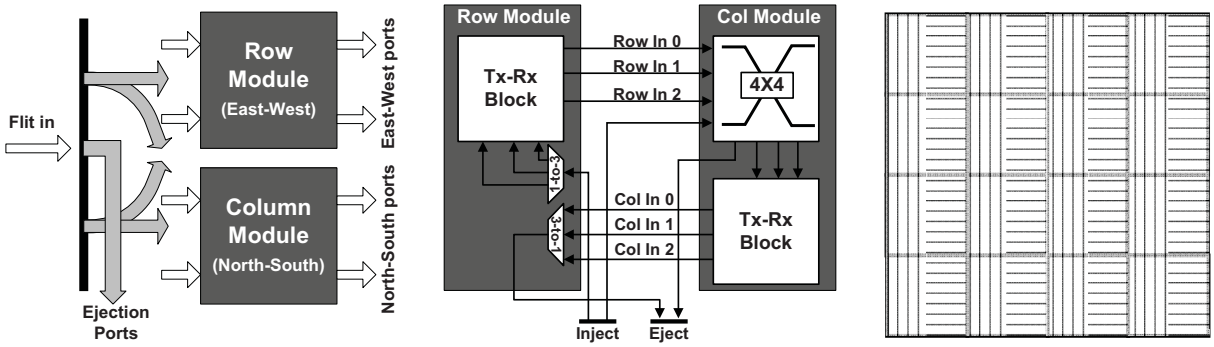


Figure 8: Free flattened butterfly: (a) concept of dimension-decomposition; (b) router architecture; (c) RGL pattern.

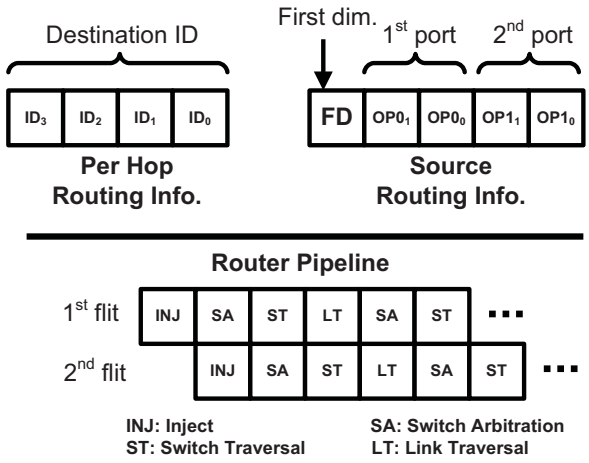


Figure 9: Up: routing information comparison for per hop routing and source routing. Down: router pipeline stages

ferred in the network. Each link employs one additional VCSEL-photodetector pair to transfer credits.

• **RGL Design:** With dimensionally-decomposed router architecture, designing the RGL pattern and align it with optical elements becomes even easier. Figure 8c illustrates one possible configuration. The RGL only needs to incorporate two types of patterns, each for one of the dimension module. Alignment is achieved as long as the patterns covers the corresponding dimension modules. Therefore, there is no need for complex RGL patterns and large alignment errors can be tolerated.

3.3 Comparison to Prior Approaches

All-to-all topologies gain favor in prior researches of optical networks, since these approaches seems to better leverage the efficiency of optical communication. In Corona NoC [15, 17], an optical crossbar is formed using multiple optical token rings to interconnect all nodes in the network. Collisions on the rings are avoided by using optical tokens to grant accesses to the rings. In Xue *et al.*'s work [25], every node in the network has FSOI links to all other nodes, and collectively these FSOI links form an optical crossbar. However, all-to-all networks scale poorly. Both Corona and Xue *et al.*'s work require a daunting number of optical devices. In Corona the large number of waveguides and micro-rings also introduces great static power consumed by the external laser source and ring trimming. In Xue *et al.*'s work, accurately controlling all FSOI links becomes an intimidating design and implementation issue. Moreover, as the bandwidth density holds constant, the large number of FSOI links also reduces the bandwidth per link and increases serialization latency.

In addition to the difficulties with the optical technology, all-to-all topologies also complicates the router design at each node. For Corona, each router has at least a wide demultiplexer to direct flits to queues on corresponding rings, and a complex arbiter to nominate a fix number of queues for optical arbitration and transmission [17]. For Xue *et al.*'s work [25], collision is handled by the exponential back-off algorithm with an initial window size of 3 and a fractional base of 1.1. It is not clear from their work how such a back-off algorithm can be implemented efficiently in the hardware. Adding to the difficulties, each of the $O(N^2)$ FSOI links needs to run a back-off algorithm independently. However, the costs of implementing the back-off algorithm is not considered in their paper.

In contrast, F^{2}BFLY trades some of the optical efficiency for better scalability, lower power consumption, and easy hardware implementation. Compared with Corona [15, 17], F^{2}BFLY has at most one more hop (3 network cycles) and one more pair of O/E and E/O conversion, but much lower static power consumption. Compared with Xue *et al.*'s work [25], it has fewer FSOI links and higher link bandwidth, and is more feasible to implement. In either case, the router of F^{2}BFLY has lower complexity and is more suitable for on-chip implementation.

4. EXPERIMENT SETUP

4.1 Methodology

We perform a comparative evaluation of F^{2}BFLY and other state-of-the-art electrical and optical networks using a cycle-accurate NoC simulator [32]. Besides synthetic traffic patterns, we also run trace-driven simulations of parallel benchmarks by interfacing the NoC simulator with an x86 CMP simulator [33]. We use ORION 2.0 [34] to estimate the power consumption of electrical routers. In addition, we use reported data in [14, 17, 18] to estimate the power for the external laser, ring modulation, and ring trimming. Fi-

nally, the power expended by VCSELs and photodetectors is estimated from the analysis presented in section 2. The network architectures evaluated are detailed in the following subsection.

4.2 Network Configuration

Table 2 lists configurations of the network architectures evaluated, including three electrical networks: mesh, flattened butterfly (FBFLY) [5], and multidrop express channels (MECS) [7]; and two optical networks: Corona and F^{2}BFLY. For each architecture, two network sizes (16-node and 64-node) are used in the evaluation.

Both FBFLY and MECS belong to the express cube family, which features high-radix routers and high wiring complexity. FBFLY is designed to leverage the rich on-chip wires and have a network diameter of 2 hops. From the topology perspective, FBFLY is exactly the same as our proposed F^{2}BFLY. However, FBFLY uses electrical links and suffers from the limited bandwidth density and power efficiency of electrical interconnects [7]. In contrast, F^{2}BFLY is built with FSOI links which provides much higher bandwidth with lower power consumption. In addition, in the original FBFLY work, the authors simply adapts a generic wormhole switching router as the high-radix router, which incurs high hardware and power overheads. In this paper, we adopt a dimension-decomposed router architecture which together with other optimizations significantly reduces hardware complexity and power consumption (see Section 3). The MECS architecture is proposed to overcome the high interconnect intensity of FBFLY. MECS has lower router radix but also more restricted connectivity, which makes it a compromise between complexity and connectivity. Corona is a channel-guided optical network. As we have discussed in Section 3.3, Corona is based on an optical crossbar formed by multiple optical token rings and has high implementation complexities and power consumption.

• **Concentration:** According to prior studies [35, 7], concentrated on-chip network provides better power-performance efficiency. In this work, we use 4-way concentration for all networks evaluated. The second and the third rows of Table 2 show the network sizes in terms of number of nodes and number of PEs. In addition, there are two kinds of concentration: *external* concentration first merges traffics of the 4 PEs and sends using a single port of the router; *internal* concentration adds more inject/eject ports to the router and directly connects the PEs to the additional ports. In general, internal concentration provides better performance but has higher hardware and power overheads. In our evaluation, we exam both internal concentration and external concentration for the electrical NoCs[1].

• **Link Bandwidth:** To perform a fair comparison, we use the iso-bandwidth analysis as that used by Grot *et al.* [7]: we keep the bi-section bandwidth constant (4,608-bit and 18,432-bit for the 16-node and the 64-node networks respectively), and calculate the link bandwidth for each of the three electrical networks. As we can see from the sixth row (link width) of Table 2, mesh has highest link bandwidth due to its lowest wiring complexity, while FBFLY has the lowest link bandwidth due to its highest wiring complexity.

[1]From our experiment results, power and performance of the two optical networks are both insensitive to concentration types.

Table 3: Configurations for the CMP

Frequency	5GHz
Processor Core	
Issue/Commit width	2
ROB size	128
L1 Cache	
Size	32KB per core
Associativity	4
Block size	64-byte
L2 Cache	
Size	1MB bank
Associativity	16
Block size	64-byte
Memory	
Size	4GB
Latency	260 cycles
Controllers	4 MCs on the corners

The results of MECS show a compromise between complexity and link bandwidth.

For Corona, we assume each waveguide carries 72 wavelengths and 4 waveguides forms a physical channel, leading to a 288-bit link. For F^{2}BFLY, by conservatively assuming that VCSELs and photodetectors have a pitch of 50um and the node area is 1.5mm×1.5mm, each FSOI link is 72-bit and 36-bit for the 16- and 64-node networks respectively. The frequencies of optical links are 10GHz and 40GHz for Corona and F^{2}BFLY respectively, conforming to the settings in [17] and [25].

• **Virtual Channels:** Only the mesh network employs multiple virtual channels, with each channel has a depth of 8-flit. For FBFLY and MECS, the depth of the input buffer is designed to account for the worst-case round-trip credit return latency. For Corona, the depth of the ejection queue is designed to cover the round-trip latency of optical tokens. Finally, since the FSOI link has a single cycle latency, the buffer size is fixed for both network sizes, also according to the credit return latency.

4.3 CMP Configuration

The CMP consists of 64 (256) PEs for networks with 16 (64) nodes. Each PE consists of an x86-like core, a private L1 cache and a shared L2 cache bank, and is connected to the corresponding router through a network interface. The whole CMP system including routers runs at 5GHz. Table 3 lists the detailed configuration for the CMP.

5. EXPERIMENT RESULTS

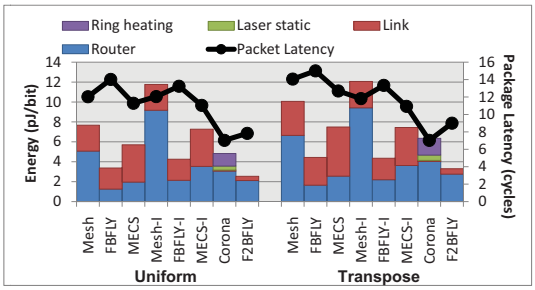
5.1 Results for Synthetic Traffics

We first study different networks with synthetic traffics. For each synthetic traffic, packets with a size of 576-bit (equivalent to a cache line transfer) are generated with the specified injection rates. The measurements are taken after the network reaches a steady state.

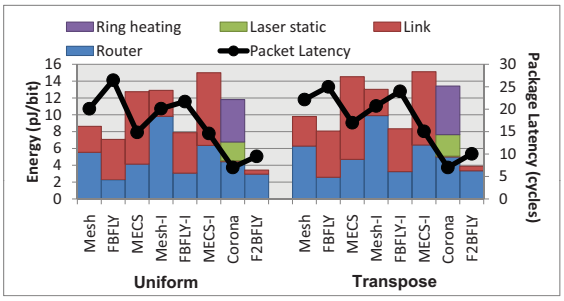
• **Performance:** Figure 10 shows the average packet latency against the injection rate for different networks. Let us first look at the results for the 16-node network. We notice that when the injection rate is low, FBFLY and MECS have comparable or slightly lower packet latencies than mesh. However, the packet latencies of both topologies rise rapidly when the inject rate increases, and the saturate throughputs are much lower than mesh. This is because, while the two high-radix networks have lower network diameter,

Table 2: Configurations of Different Network Architectures. The "-I" versions of networks refer to the cases with internal concentration.

		Mesh (-I)		FBFLY (-I)		MECS (-I)		Corona		F^2BFLY	
Network	Number of nodes	16	64	16	64	16	64	16	64	16	64
	Number of PEs	64	256	64	256	64	256	64	256	64	256
	Concentration degree	4	4	4	4	4	4	4	4	4	4
	Network diameter	6	14	2	2	2	2	1	1	2	2
	Link width	576	576	144	72	288	288	288	288	72	36
Router	Input ports	5 (8)	5 (8)	7 (10)	15 (18)	7 (10)	15 (18)	1	1	7	15
	Output ports	5 (8)	5 (8)	7 (10)	15 (18)	3 (6)	3 (6)	15	63	7	15
	Switching fabrics	5×5 (8×8)	5×5 (8×8)	7×7 (10×10)	15×15 (18×18)	7×3 (10×6)	15×3 (18×6)	1-to-15	1-to-63	4×4 & 1-to-3 & 3-to-1	8×8 & 1-to-7 & 7-to-1
	VC per port	5	5	1	1	1	1	1	1	1	1
	VC depth	8	8	10	15	10	15	8	16	8	8

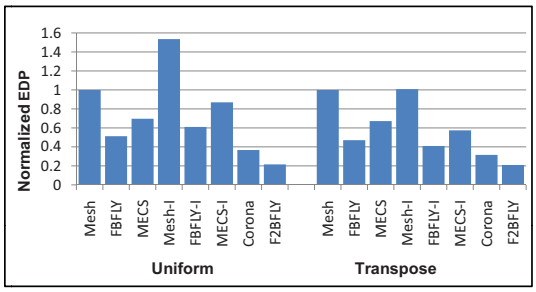


(a) 16-node network

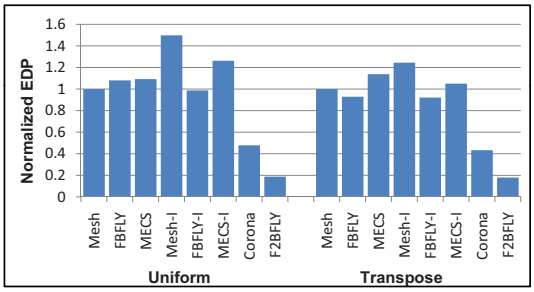


(b) 64-node network

Figure 11: Energy consumption for uniform and transpose traffics, when the injection rate is 0.05 packets/cycle/node.



(a) 16-node network



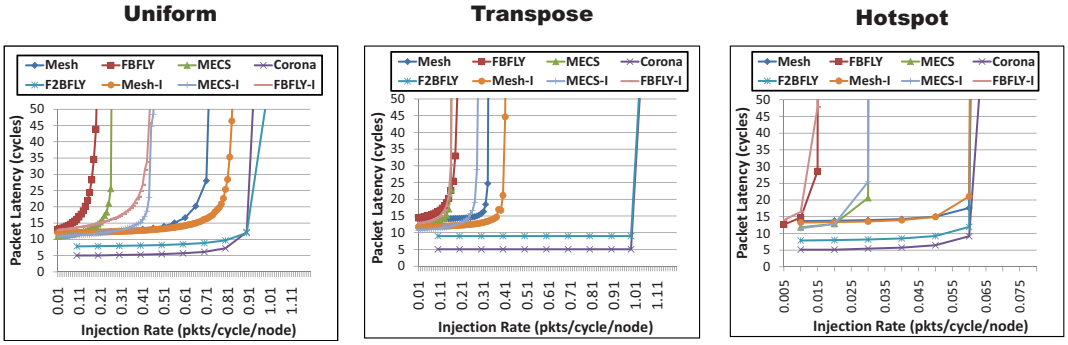
(b) 64-node network

Figure 12: Energy-delay products calculated from results in Figure 11.

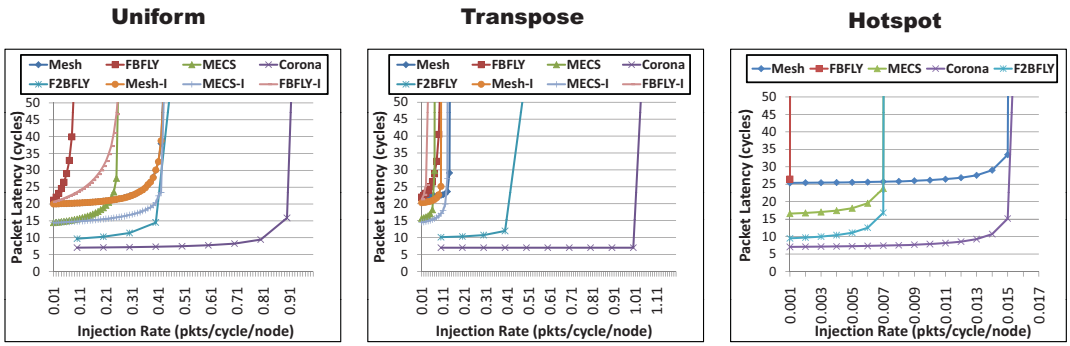
the lower link bandwidth of them results in a longer serialization latency, which will cause the network to saturate early. On the other hand, both optical networks have consistently lower packet latencies than all the electrical networks, thanks to the significantly higher bandwidth of the optical links. It is worth noting that due to the higher hop counts, F^2BFLY has slightly higher packet latencies than Corona. However, both networks have roughly the same saturation throughputs. Moreover, we also see that internal concentration improves the performances for the electrical networks for uniform and transpose traffics, due to the increased injection and ejection bandwidth. The performances for the hotspot traffic see almost no improvement, because in the hotspot traffic the performance is restricted by the single destination instead of the injection/ejection ports.

In the 64-node network, the packet latencies of MECS at low loads are even lower than mesh, due to the increased network scale. In contrast, the performance of FBFLY becomes worse, exhibiting largest packet latencies among all networks for all loads. The primary cause for this phenomenon is the halved link bandwidth for FBFLY when the network size increases from 16-node to 64-node. This implies that FBFLY has inferior scalability than other networks. In the 64-node network, the two optical architectures continue to perform better than electrical networks. However, we notice that F^2BFLY now has significantly lower saturation throughput than Corona. This is because we conservatively halved the link bandwidth of F2BFLY for the 64-node network, but optimistically keep the link bandwidth for Corona. Finally, internal concentration continue to improve the performances for electrical networks for uniform and transpose traffics. The results with internal concentration for the hotspot traffic are not shown, since internal concentration does not improve the performance in this case.

•**Power:** We set the injection rate to 0.05 packets/cycle/node and measure the energy consumption for the uniform and the transpose traffics. Figure 11 shows the results for all architectures for both 16- and 64-node networks. The black curve in each plot shows the average packet latencies. For the 16-node network, the F^2BFLY has the lowest energy consumption. While Corona has slightly lower packet



(a) 16-node network



(b) 64-node network

Figure 10: Performance results for synthetic traffics. The "-I" versions of networks refer to the cases with internal concentration. For the 64-node network, the results with internal concentration for the hotspot traffic are not shown, since we already show with the 64-node network it does not improve performance for the hotspot traffic.

latency than F^{2}BFLY, it has significantly higher energy consumption due to the more complex router architecture and the static power consumed by the external laser source and ring trimming. For the 64-node network, we see a rapid increase of power spent by the laser source and ring trimming: the static energy accounts for more than 60% of total energy consumption of Corona, largely offsetting the power efficiency.This result implies that when the network size goes up, the static energy waste impedes the energy efficiency of Corona. Finally, we also notice that internal concentration results in slightly lower packet latency but also higher energy consumption.

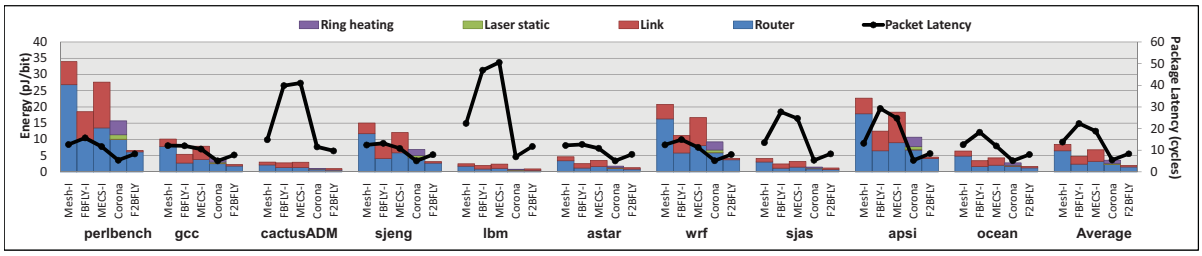
For a more comprehensive comparison, we calculate the energy-delay products (EDP) from results in Figure 11 and plot them in Figure 12. As we can see, in the 16-node network, the two optical networks have the lowest EDP. In addition, the EDP of F^{2}BFLY is 41% lower than that of Corona. In the 64-node network, while the optical networks continue to have the lowest EDP, Corona's EDP increases significantly due to the large static power consumption. In this case, F^{2}BFLY's EDP is 80% lower than Corona's EDP. In addition, we see that the EDP results with and without internal concentration are roughly the same, because internal concentration slightly decreases the packet latency, but increases the power consumption.

5.2 Results for Applications

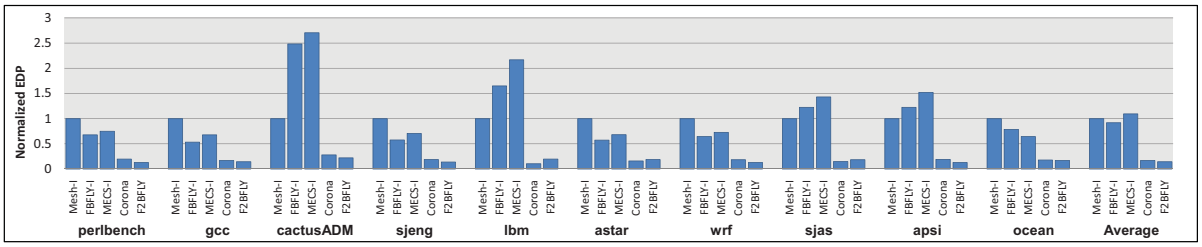
We further evaluate networks by simulating a CMP running different applications. To make the simulation time tractable, we only simulate a 16-node network with 64 processing elements. Furthermore, for the electrical networks we only evaluate internal concentration since our studies above show that internal concentration results in slightly better performance and roughly the same EDP. The network and CMP configurations are detailed in Table 2 and 3. We collect traces from a diverse set of 33 benchmarks including SPEC CPU2006 benchmarks, applications from SPLASH-2 and SPEC OMP benchmark suites, and four commercial workloads (`sap`, `sjas`, `sjbb`, `tpcw`). In addition, we choose representative execution phases using [36] for all the workloads.

Figure 13 shows the results for network performance and power when running the applications. Due to space limit, we only show results for 10 representative applications out of the 33 benchmarks. The last set of bars in each figure show the average results for all 33 benchmarks. From the results, we first notice that the two optical networks have the lowest packet latencies and power consumptions across all benchmarks. Second, the high-radix electrical networks (FBFLY and MECS) have slightly lower latencies than mesh for some benchmarks (`gcc`, `astar`), but in general higher latencies than mesh for most benchmarks. A closer look at the benchmarks reveals that `gcc` and `astar` have quite low injection rates (less than 0.02 flits/cycle/node), however on average the injection rates of all benchmarks are high (around 0.1 flits/cycle/node) partly because of the 4-way concentration used.

Figure 13b shows the normalized energy-delay products. We notice that the EDPs for the high-radix electrical networks (FBFLY and MECS) are lower than mesh for some



(a) Packet latency and energy consumption



(b) EDP

Figure 13: Network performance and energy results for applications.

benchmarks, but higher for others. On average, the EDP for FBFLY is 10% lower and the EDP for MECS is 10% higher compared with mesh. Our proposed F^{2}BFLY is shown to have the lowest EDP, which is 15% better than Corona. While the saving in EDP is not as significant as with the synthetic traffics shown above, we believe the energy advantage of F^{2}BFLY will increase as the network size scales up as predicted by the synthetic traffic results.

6. RELATED WORK

Network-on-chip (NoC) has gained interests from researchers due to the high on-chip communication demands of contemporary and future many-core architectures. Conventional research efforts are focused on router architectures with lower router latencies [2, 3, 4] and efficient topologies [5, 6, 7, 8]. The seminal work by Peh *et al.* [2] establishes a low latency wormhole switching router architecture that is widely adopted by other researchers. Both Kim *et al.* [5] and Grot *et al.* [7] propose to use high-radix networks to better exploit the on-chip wiring resources. Dimension-decomposition is proposed by Kim *et al.* [30] to further simplify router architectures. Our proposed router architecture for F^{2}BFLY inherits the basic organization of the wormhole switching router proposed by Peh *et al.*'s. In addition, we apply dimension-decomposition to reduce the hardware complexity of the high-radix router, which results in lower hardware and power overhead than the high-radix routers used in prior work [5, 7]. While in terms of topology F^{2}BFLY is the same as FBFLY [5], we use both FSOI interconnects and an optimized router architecture to tackle the wiring and the hardware complexities.

To overcome the intrinsic inefficiency of electrical interconnects, researchers have explored optical NoCs based on emerging on-chip photonic technologies. Prior work [12, 13, 14, 15, 16, 17, 18] mostly uses channel-guided optical interconnects, which suffers from significant static power consumption and poor scalability. Most recently, Xue *et al.* [25] present an optical network based on FSOI links. However, in their approach micro-mirrors are used to guide laser beams and an all-to-all architecture is proposed using large amount of optical elements. The resulted system is not amenable for on-chip implementation and has low scalability. In contrast, our proposed F^{2}BFLY is based on FSOI technologies with mature fabrication processes. In addition, in designing the network architecture, we trade slight performance reduction for lower hardware complexity and power consumption, resulting in an efficient optical network using FSOI links.

7. CONCLUSION

In this work, we present F^{2}BFLY network-on-chip which is based on novel FSOI technologies. Compared to state-of-the-art optical NoCs, it has lower hardware complexity and power consumption, and also better scalability. We evaluate F^{2}BFLY together with recently developed electrical and optical NoCs, and our results show that F^{2}BFLY has lowest power consumption and energy-delay products for both synthetic traffics and benchmarks. In addition, when the network scale grows up energy efficiency of F^{2}BFLY is even better than Corona, a representative channel-guided optical network. These results from comparative evaluation show that F^{2}BFLY is a potential solution for future efficient on-chip interconnection substrates.

8. REFERENCES

[1] J. Y. Duato, S. Yalamanchili, and L. Ni. *Interconnection Networks: An Engineering Approach.* Morgan Kaufman, 2003.

[2] L. Peh and W. J. Dally. A delay model and speculative architecture for pipelined routers. In *Proc. of the 7th Intl. Symp. on High Performance Computer Architecture*, page 255, 2001.

[3] L. Peh and W. J. Dally. Flit-reservation flow control. *IEEE Trans. on Parallel and Distributed Systems*, 3(3):194–205, 2000.

[4] M. Ahn and E.J. Kim. Pseudo-circuit: accelerating communication for on-chip interconnection networks. In *Proc. of the 43th Intl. Symp. on Microarchitecture*, pages 399–408, 2010.

[5] J. Kim, J. Balfour, and W. Dally. Flattened butterfly topology for on-chip networks. In *Proc. of the 40th Intl. Symp. on Microarchitecture*, pages 172–182, 2007.

[6] B. Grot and S.W. Keckler. Scalable on-chip interconnect topologies. In *Workshop on Chip Multiprocessors Memory Systems and Interconnects*, 2008.

[7] B. Grot et al. Express cube topologies for on-chip interconnects. In *Proc. of the 15th Intl. Symp. on High Performance Computer Architecture*, pages 163 –174, 2009.

[8] A. Kumar et al. Express virtual channels: towards the ideal interconnection fabric. In *Proc. of Intl. Symp. on Computer Architecture*, pages 150–161, 2007.

[9] 2007 International technology roadmap for semiconductors, at `http://www.itrs.net`

[10] L. Cheng et al. Interconnect-aware coherence protocols for chip multiprocessors. *Prof. of the 33rd Intl. Symp. on Computer Architecture*, 34(2):339–351, 2006.

[11] N. Magen et al. Interconnect-power dissipation in a microprocessor. In *Proc. of SLIP '04*, pages 7–13, 2004.

[12] N. Kirman et al. Leveraging optical technology in future bus-based chip multiprocessors. In *Proc. of the 39th Intl. Symp. on Microarchitecture*, pages 492–503, 2006.

[13] A. Shacham, K. Bergman, and L. P. Carloni. On the design of a photonic network-on-chip. In *Proc. of the First Intl. Symp. on Networks-on-Chip*, pages 53–64, 2007.

[14] C. Batten et al. Building manycore processor-to-dram networks with monolithic silicon photonics. In *Proc. of the 16th IEEE Symp. on High Performance Interconnects*, pages 21–30, 2008.

[15] D. Vantrease et al. Corona: System implications of emerging nanophotonic technology. In *Proc. of the 35th Intl. Symp. on Computer Architecture*, volume 36, pages 153–164. ACM, 2008.

[16] Y. Pan et al. Firefly: illuminating future network-on-chip with nanophotonics. 37(3):429–440, 2009.

[17] D. Vantrease et al. Light speed arbitration and flow control for nanophotonic interconnects. In *Proc. of the 42nd Intl. Symp. on Microarchitecture*, pages 304–315, 2009.

[18] X. Zhang and A. Louri. A multilayer nanophotonic interconnection network for on-chip many-core communications. In *Proc. of Design Automation Conference*, pages 657–660, 2010.

[19] M. Haurylau et al. On-chip optical interconnect roadmap: Challenges and critical directions. *IEEE Journal of Selected Topics in Quantum Electronics*, 12(6):1699 –1705, 2006.

[20] Y. Pan, J. Kim, and G. Memik. Flexishare: Channel sharing for an energy-efficient nanophotonic crossbar. In *Proc. of the 16th Intl. Symp. on High Performance Computer Architecture*, pages 1–12, 2010.

[21] A. Imamura, V. Karagodsky, B. Pesala, F. Koyama, and C.J. Chang-Hasnain. Multi-wavelength VCSEL array based on high contrast sub-wavelength grating. In *Proc. of LEOS Annual Meeting Conf.*, pages 811 –812, 2009.

[22] F. Koyama and M. Arai. Long-wavelength GaInAs/GaAs VCSELs and multiwavelength arrays. In *Photonics: Design, Technology, and Packaging*, volume 5277, pages 146–154, 2004.

[23] V. Karagodsky et al. Monolithically integrated multi-wavelength VCSEL arrays using high-contrast gratings. *Opt. Express*, 18(2):694–699, Jan 2010.

[24] B. E. Lemoff, D. Babic, and P. Schneider. Monolithic multiple wavelength VCSEL array, US Patent No. 6117699, 2000.

[25] J. Xue et al. An intra-chip free-space optical interconnect. In *Proc. of the 37th Intl. Symp. on Computer Architecture*, pages 94–105, 2010.

[26] J. A. Arns, W. S. Colburn, and S. C. Barden. Volume phase grating for spectroscopy, ultrafast laser compressors, and wavelength division multiplexing. In *Proc. SPIE*, volume 3799, 1999.

[27] P.-A. Blanche et al. Volume phase holographic gratings: large size and high diffraction efficiency. In *Opt. Eng.*, volume 43, 2004.

[28] P. Westbergh et al. 850nm VCSEL operating error-free at 40 Gbit/s. In *Proc. 22nd IEEE International Semiconductor Laser Conference*, pages 154 –155, 2010.

[29] A. Ramam, G. K. Chowdhury, and S. J. Chua. An approach to the design of highly selective resonant-cavity-enhanced photodetectors. *Applied Physics Letters*, 86(17):171104 –171104–3, 2005.

[30] J. Kim et al. A gracefully degrading and energy-efficient modular router architecture for on-chip networks. In *Proc. of the 33rd Intl Symp. on Computer Architecture*, pages 4–15, 2006.

[31] J. Kim et al. A novel dimensionally-decomposed router for on-chip communication in 3D architectures. In *Proc. of the 34th Intl Symp. on Computer Architecture*, pages 138–149, 2007.

[32] J. Kim et al. Design and analysis of an NoC architecture from performance, reliability and energy perspective. In *Proc. of the 2005 ACM Symp. on Architecture for Networking and Communications Systems*, pages 173–182, 2005.

[33] R. Das et al. Aérgia: exploiting packet latency slack in on-chip networks. In *Proc. of the 37th Intl. Symp. on Computer Architecture*, pages 106–116, 2010.

[34] A. B. Kahng et al. ORION 2.0: a fast and accurate NoC power and area model for early-stage design space exploration. In *Proc. of the Conference on Design, Automation and Test in Europe*, pages 423–428, 2009.

[35] J. Balfour and W. J. Dally. Design tradeoffs for tiled CMP on-chip networks. In *Proc. of the 20th Intl. Conf. on Supercomputing*, pages 187–198, 2006.

[36] H. Patil et al. Pinpointing representative portions of large Intel® Itanium® programs with dynamic instrumentation. In *Proc. of the 37th Intl. Symp. on Microarchitecture*, pages 81–92, 2004.