This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

UNION: A Unified Inter/Intrachip Optical Network for Chip Multiprocessors

Xiaowen Wu, *Student Member, IEEE*, Yaoyao Ye, *Student Member, IEEE*, Jiang Xu, *Member, IEEE*, Wei Zhang, *Member, IEEE*, Weichen Liu, *Member, IEEE*, Mahdi Nikdast, *Student Member, IEEE*, and Xuan Wang, *Student Member, IEEE*

***Abstract*— As modern computing systems become increasingly complex, communication efficiency among and inside chips has become as important as the computation speeds of individual processing cores. Traditionally, to maximize design flexibility, interchip and intrachip communication architectures are separately designed under different constraints. Jointly designing communication architectures for both interchip and intrachip communication could, however, potentially yield better solutions. In this paper, we present a unified inter/intrachip optical network, called UNION, for chip multiprocessors (CMPs). UNION is based on recent progresses in nanophotonic technologies. It connects not only cores on a single CMP, but also multiple CMPs in a system. UNION employs a hierarchical optical network to separate interchip communication traffic from intrachip communication traffic. It fully utilizes a single optical network to transmit both payload and control packets. The network controller on each CMP not only manages intrachip communications, but also collaborates with each other to facilitate interchip communications. We compared UNION with a matched electrical counterpart in 45-nm process. Simulation results for eight real CMP applications show that on average UNION improves CMP performance by 3× while reducing 88% of network energy consumption.**

***Index Terms*— Chip multiprocessor (CMP), interchip optical network, optical NoC, photonic interconnects.**

I. Introduction

MODERN computing systems have become increasingly complex to satisfy the growing performance demanded by applications. As the number of transistors available on a single chip increases to billions, chip multiprocessor (CMP) has become an attractive platform delivering high performance with limited power budget. In a complex CMP system, the communication efficiency among and within chips is very important for the overall system performance. If the data-access latency becomes too large, the processing cores would spend much of their time simply waiting, wasting the processing power.

Manuscript received April 1, 2012; revised December 4, 2012 and March 21, 2013; accepted May 9, 2013. This work was supported in part by RGC DAG, HKUST RPC, and the Nanoscience and Nanotechnology Program. *(Corresponding author: J. Xu.)*

X. Wu, Y. Ye, J. Xu, W. Liu, M. Nikdast, and X. Wang are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science of Technology, Hong Kong (e-mail: wxxaf@ust.hk; yeyaoyao@ust.hk; jiang.xu@ust.hk; weichen@ust.hk; mnikdast@ust.hk; eexwang@ust.hk).

W. Zhang is with the School of Computer Engineering, Nanyang Technological University, 637820 Singapore (e-mail: zhangwei@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2013.2263397

For interchip communications, bus-based and ad-hoc architectures are still popular, and signals are mostly transmitted by electrical interconnects on printed circuit boards (PCB). The limitations of electrical interconnects such as high-delay and high-power consumption, are already shown in high-performance systems. For intrachip communication, architectures gradually move from ad-hoc and bus-based architectures to network-on-chip (NoC) to alleviate issues including poor scalability and limited bandwidth [1], [2]. As semiconductor technologies continually scale feature size down and more cores are integrated on the chip, conventional metallic interconnects have become the bottleneck of NoC. They consume large portion of the on-chip power. In addition, for relative long interconnect, it does not scale down as the gate. To keep the delay constant, repeater insertion and wire resizing are necessary, which will not only consume much more power, but also reduce the bandwidth density.

Optical interconnects, with advantages including ultrahigh throughput, low-delay and low-power consumption, are proposed to replace both inter- and intrachip electrical wires. For interchip communication, optical interconnects are studied for more than a decade and many promising research results are proposed [3], [4]. For intrachip communication, with silicon photonics being mature, optical links are suggested to replace long metal wires on chip [5]–[8].

Traditionally, interchip and intrachip communication architectures are separately designed. This is because there is a huge performance gap between intra- and interchip electrical interconnects. First, the delay of on-chip wires are much smaller than the off-chip ones because of the substantially smaller physical length, resistance, and capacitance. Second, the limited number of I/O would severely restrict the possible off-chip bandwidth while the on-chip wires are much more abundant. Third, large drive power is required for the off-chip wire which is with high capacitance. Crosstalk issue further limits the possible bandwidth of the off-chip interconnect. All these make intrachip and interchip interconnects mismatched, and the two architectures are always designed with different properties and protocols.

For optical interconnects, unlike electrical wires, the inter- and intrachip channels can be interconnected seamlessly. Both on-chip and off-chip channels can be implemented with optical waveguides and they can be interconnected with passive couplers. The allowed operating bandwidths of both waveguides are broad enough for real applications. Given the high

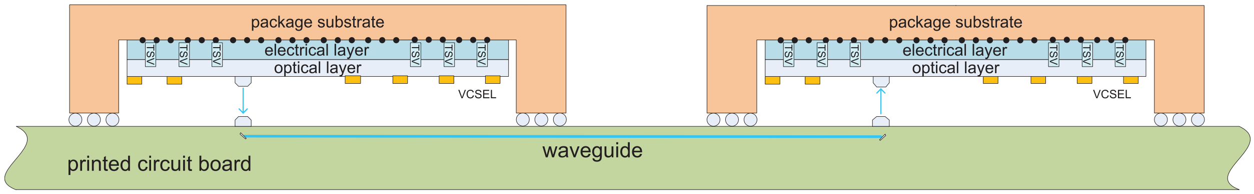


Fig. 1. UNION cross-sectiontional overview (not to scale). Optical layer and electrical layer are back-to-back stacked with TSVs. Each integrated chip is surface mounted to a PCB on which polymer waveguides are fabricated. Lens and mirrors are incorporated on both waveguides on chip and on board to form a free space coupler.

propagation speed of light, the delay difference for on-chip and off-chip links is minor. In addition, if the small transmission loss is neglected, the transmission power is independent of transmission length. All these make the optical interchip and intrachip interconnects well matched, and a unified design becomes natural. Motivated by these observations, we propose a unified inter/intrachip optical network, called UNION.

The cross-sectiontional view of UNION is shown in the Fig. 1. Multiple chips are integrated on the PCB. In each chip, an optical die is stacked with the processor die with 3-D integration technology. On the optical die, the photonic devices are fabricated, which include waveguides, switches, and photodiodes. VCSELs are bonded on the chip to supply optical power [9], [10]. All the photonic devices can be accessed by the processor die with through-silicon-vias (TSVs). For interchip communication, polymer waveguides are fabricated on a PCB as transmission medium. Integrated chips are directly surface mounted to a PCB by a conventional ball grid array solder process similar to [3]. Mirrors and lens arrays are incorporated to couple the light between silicon waveguides on-chip and the polymer waveguides on board. The chips can optically communicate with each other as shown in Fig. 1.

The characteristics of UNION are as follows. First, it is an all-optical network that, data can not only be transmitted optically among cores on the same chip, but also be optically transmitted among cores on different chips. There is no electrical-to-optical (EO)/optical-to-electrical (OE) conversion between the intrachip and interchip optical networks, saving electrical buffers that consume large silicon resources and also power. Second, for on-chip subnetwork, we propose a central controller to arbitrate all transaction requests wisely with global information while it also saves path setup time. In this design, a single optical network is used to transmit both payload packets and control packets. Third, we consider the proper network floorplan and power control mechanism to reduce the power consumption as well as crosstalk.

II. Related Work

Optical interconnects for chip level communication are proposed for more than a decade. Polymer waveguides on board [11], [12], fibers [13] and free space [14], [15] are proposed as mediums for light transmission. Among these techniques, the polymer waveguide fabricated on PCB is especially favored for its compatibility with PCB design process. Another feature is the possibility to integrate splitters and combiners that are useful for buslike structures [16]. In UNION, polymer waveguides are implemented as buses for interchip communication.

In most of the proposed chip-level designs, optical interconnection is point-to-point and there is an extra sender/receiver chip responsible for sending/receiving data. With advanced nanophotonics on chip, the extra sender/receiver chips can be omitted and more scalable networks become feasible. Batten *et al.* [17] proposed a processor-to-DRAM network without the extra sender/receiver chips. Processors are grouped together by electrical networks on a chip, and they can access off-chip memory through optical channels. An extra switching chip is employed to switch optical channels between core groups and DRAMs. Koka *et al.* [18] proposed a silicon-photonic network to enable a scalable power-efficient system with multiple chips integrated in a single package. In UNION, extra sender/receiver chips are also omitted. Our design distributes the chips with more distances such that the thermal density can be reduced.

As nanophotonics become more and more mature, optical interconnects are considered to replace on-chip electrical wires. Kapur *et al.* [19] compared the optical interconnects and electrical wires driven by repeaters from both delay and power perspectives. Results show that optical interconnects are favored in global communication. Beausoleil *et al.* [20] examined the potential of replacing the global electronic interconnects of future chips with photonic interconnects in.

With the silicon photonic technologies, different on-chip network architectures are proposed. Kirman *et al.* [5] presented an optoelectrical hierarchical bus for future CMPs with cache-coherence supported. An optical loop at the top is for global communication and the bottom electrical wires are used for local interconnects. Pasricha *et al.* [21] proposed an optical ring waveguide to replace global pipelined electrical interconnects while preserving the interface with bus protocol standards such as AMBA AXI. O'Connor presented a full connected optical NoC based on the special λ-router with wavelength division multiplexing (WDM) technology [22]. Shacham *et al.* [7] proposed an hybrid optical NoC. It combines an optical circuit-switched network with an electrical packet-switched network. Electrical network is used for path setup and short packet transmission. Joshi *et al.* [23] presented a photonic Clos network. Long electrical links between routers are replaced by optical ones, which provides more uniform latency and higher throughput compared with

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

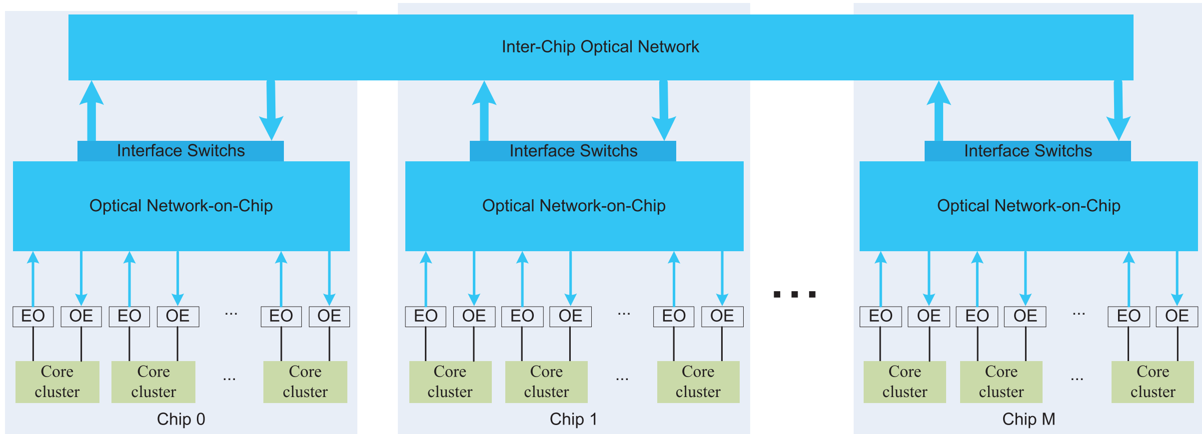


Fig. 2. UNION architecture overview. Optical NoCs are responsible for on-chip communication. A separate interchip optical network interconnecting all NoCs helps to build an all-optical inter/intrachip network.

mesh network. Cianchetti *et al.* [8] proposed a packet-switched optical network. The packet may pass through multiple routers without being buffered as long as no collision happens. Electrical buffer is implemented to buffer the packet if the collision happens. Vantrease *et al.* [6] proposed Corona architecture that uses optical interconnects for both intercore communication and off-stack communication to memory. Cores are fully interconnected with a photonic crossbar. A distributed optical token-based arbitration scheme is proposed for channel allocation. Pan *et al.* [24] proposed Firefly architecture as a hybrid hierarchical on-chip network. It implements an electrical interconnect for short distance transmission and an optical crossbar for long distance transmission. The crossbar is partitioned into smaller crossbars with localized arbitration. FlexiShare proposed by Pan *et al.* [25] implemented a flexible optical crossbar in which each data channel is accessible for all cores to write and read. Special token stream arbitration protocol is proposed to cope with the flexibility. In our UNION, an optical intrachip subnetwork is proposed to address the on-chip communication. It is a fat tree with centralized control protocol, and it is codesigned with interchip subnetwork.

This paper is a substantial extension of our previous paper [26] with the following major additions. First, the arbitration scheme for interchip network and the hardware implementation of network controller are detailed. More discussion of proposed routing algorithm is also provided. Second, we redesign the optical router and saved microresonators (MRs). Floorplan is also given to minimize the optical power loss. Third, we extend our simulations to evaluate the scalability of the system, and more accurate power model is used in power analysis. Finally, the on-chip subnetwork of UNION is also compared with the related works to show the design tradeoffs.

III. UNION Architecture

An overview of the UNION architecture is shown in Fig. 2. It includes an interchip optical network and multiple intrachip optical networks. While intrachip communications are handled by each optical NoC independently, interchip communications require the collaboration of multiple optical NoCs on different chips through the interchip network. Optical NoCs are connected to the interchip network through interface switches. Each chip has a network controller. The network controllers not only manage the intrachip networks, but also collaborate with each other to facilitate interchip communications. In UNION, long electrical interconnects are avoided, and there is no power-hungry OE or EO conversions on the paths. In the following, we detail the intrachip and the interchip optical networks along with the network protocols.

A. Intrachip Network

UNION uses a hierarchical optical NoC (as shown in Fig. 3) for each chip. The optical routers are connected in fat tree topology, and both payload and control data is transmitted in this single fat tree network. Fat tree is widely adopted in high-performance systems [27], [28], and also NoC designs [29], [30]. In UNION, it is required that an on-chip network can be integrated with interchip network in a hierarchical manner, and fat tree especially satisfies this requirement with inherent hierarchical property. It is nature to extend the top-level routers on the tree to interconnect the off-chip network, forming a larger system. Fat tree is also suitable for central-control protocol proposed that helps setup the interchip network and also boosts the performance of intrachip network. Besides the data subnetwork, the control subnetwork can also be implemented as fat tree topology, which provides the opportunity of integrating two subnetworks together as we proposed here. To highlight these merits of the fat tree topology, we compare it with mesh topology. The mesh topology does not own the same hierarchical property as fat tree. If the routers at the network edge are connected to off-chip network, these routers would be the bottleneck as they have to support more traffics. In addition, the topology of control subnetwork with the central controller would not be a mesh and thus different from the data subnetwork, implying that they cannot be combined together. In addition, the control network may cross with the data network, introducing many waveguide crossings.

In our fat tree, each router connects two parent routers through upward links and two children routers through downward links. The top-level routers are connected to the interchip

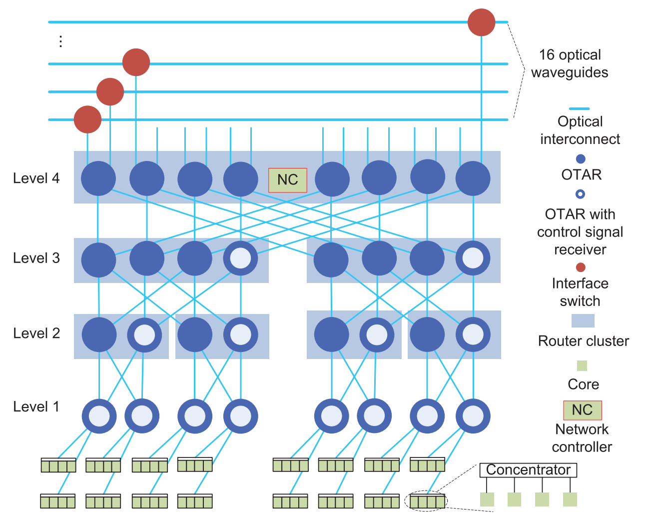
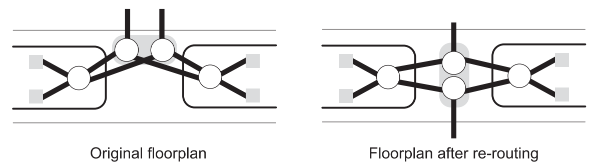
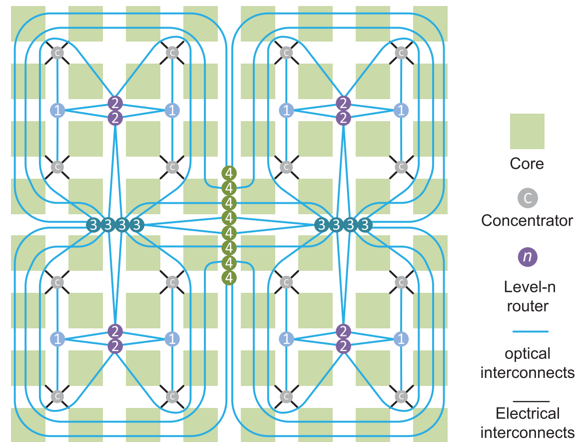


Fig. 3. Intrachip optical network. Top parallel waveguides are actually parts of interchip networks. The network controller residing on the top level of tree is responsible for the whole network configuration.



(a) Illustration of waveguide re-routing



(b) Final floorplan with most of crossings avoided

Fig. 4. Floorplan of fat tree.

optical network by interface switches, and the leaf routers are connected to core clusters. A core cluster is composed of a concentrator with four cores. The cores within a concentrator communicate with each other through a local electrical crossbar. Cores can also send data out to or receive data from optical network through the EO and OE interfaces at concentrator. This hybrid approach considers both advantages of power and performance in short electrical link and long optical link. The optical routers are controlled by the network controller residing at the top of tree. We group the routers that are physically close to each other as router clusters, and each cluster as a whole would receive the control information from the network controller.

From Fig. 3, there are many waveguide crossings. We assume there is only single layer of silicon waveguides because fabricating more than one layer is too demanding in the near future. As the network size grows, the number of crossings increases quickly. When the light passes the crossing, some power would be leaked to other waveguides. This would introduce power loss on the path, which would in turn augment the power requirement on the laser. In addition, the leaked optical power would be a noise source to other optical signals, introducing crosstalk issues that may severely confine the network size [31]. To alleviate the problem, we propose a new floorplan. In traditional floorplan, H-tree is always used with the same number of crossing as shown in Fig. 3. To reduce the crossings, we can reroute the waveguides such that some waveguides can circle around to prevent from crossing with each other. From Fig. 4(a), the crossings can be avoided by selecting another direction when we route the waveguides from level-1 routers to level-2 routers. The Fig. 4(b) shows the final floorplan of our fat tree. The number of crossing is significantly reduced. Although the increased waveguide may introduce extra power loss, the reduced crossing loss would be more significant, especially when the network size is large. With the parameters described in Section IV, it shows that 16% optical power is saved by rerouting the waveguides.

1) Routing Protocol: In UNION, if both sides of a transaction are within the same concentrator, packets are transmitted through a local electrical crossbar. On the other hand, if a core needs to send a packet out of the concentrator, it first tries to reserve an optical path to the destination concentrator. If the path is reserved successfully, the packet would be transmitted optically to the destination concentrator that would send it to the destined core through the local electrical crossbar.

In traditional optical circuit switching networks, a separate electrical control network is required for path maintenance [7]. The control packets are transmitted by this control network. They can also be sent in the optical network but with EO/OE conversions at every router along the path [32]. In both cases, the path setup involves multiple hops and thereby introducing large latency. Different from the above methods, we implement a special central-control unit called as the network controller to configure all routers. In particular, all concentrators and clusters are optically interconnected to the network controller that sends/receives control information to/from them. The transmission of control information can be served by the original data network given that the control network topology can also be implemented as a tree. The control information and payload data information would be sent in different wavelengths to avoid collisions. Therefore, the costs of the control network are only some control signal transceivers at both ends of the links. The potential disadvantage of the central control is the delay of control information transmission, which is surely problematic for electrical wires. In the optical domain, it is, however, no longer a problem. In UNION, on-chip transmission delay can be contained within one clock cycle. The network controller is virtually near to all routers, and it can provide global, instead of local, network information to them quickly.

Algorithm 1 Requests Arbitration

```
Require: Requests, Link_state
Select n requests and store them into R;
for All requests in R do
    P[i] ⇐ Find_path(R[i]);
end for
Temp_link_state ⇐ NULL;
for All paths in P do
    if Check_collision(P[i],Link_state)==false &&
    Check_collision(P[i],Temp_link_state)==false then
        Temp_link_state ⇐ Temp_link_state ∪ P[i];
        Granted_group.add(P[i]);
    end if
end for
Link_state ⇐ Link_state ∪ Temp_link_state;
return Granted_group;
```

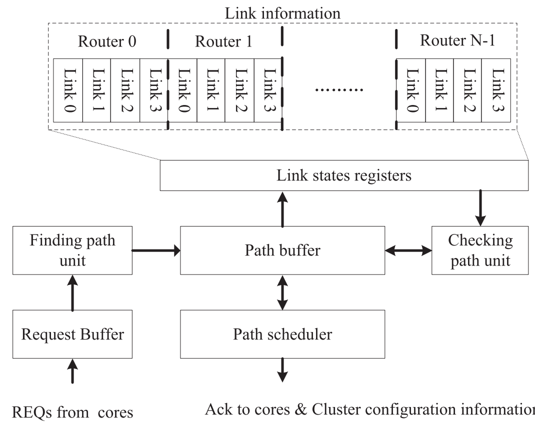


Fig. 5. Network controller structure. Only the components for on-chip communication are shown.

With the network controller, the routing protocol is pretty straightforward. If a concentrator wants to start a transmission, it would send a request to the network controller that would reserve the path if possible and then grant the request. Once the source concentrator receives a grant signal, it can send out data propagating along the reserved path. After the transmission is finished, a tear down signal will be sent from the source core to the network controller to ask for path release. In the whole process, only a limited number of control signals need to be transmitted. Compared with distributed path setup mechanisms, UNION can significantly reduce the collisions with global information and thereby improving the network performance.

2) Network Controller: The network controller on each chip, located at the top of fat tree, is responsible for requests arbitration and path configuration. They not only control on-chip networks, but also work in close cooperation with each other for interchip communications. Here, we only introduce part of the network controller for on-chip communication. The arbitration algorithm is from Algorithm 1. Initially, we select the candidate requests from the request buffer. Then, we decide the path for each request according to the routing algorithm discussed later. After the path selection, we make sure that the selected path would not collide with the existing paths by checking the states of the links. In addition, we make sure that the selected paths would not collide with each other. Finally, we update the link states of the network based on newly added paths and return the granted requests. Assuming the number of cluster is n, the complexity of this algorithm is $O(n \log n)$, given that it checks n paths and the longest path of a request is composed of $2 \log_2 n$ links.

The design of the network controller is shown in Fig. 5. The request buffer unit is responsible for receiving requests from cores. The path finding unit would fetch the requests from the request buffer, determine a path for each request according to a routing algorithm, and then store the path information to path buffer. The finding path processes can run in parallel for all requests as they are independent with each other. All the link information of the network is stored in special registers indexed by link ID. Each register is with 1 representing the corresponding link is busy or with 0 if the link is free. After the path is determined, the availability of the path is decided by checking all links on the path. Specifically, simple AND gates are implemented to decide whether the path is available (with an outcome 1) or not (with an outcome 0). The checking path process is parallel for all candidate requests. As there are many requests being processed simultaneously, a path scheduler is implemented to select a set of nonoverlapping requests. The requests can be compared with each other in parallel to quickly decide the collision sates and find the nonoverlapping set. After the selection of successful paths, link information is updated and path configuration information is sent out to corresponding clusters. The clusters in turn would configure the related MRs to setup the paths for selected requests. In the following sections, we will first discuss the routing algorithm and then optical router design.

3) Routing Algorithm: For the fat tree network, a typical minimum-path routing algorithm is the turnaround routing algorithm. Specifically, a packet is routed upward from the source core until it reaches a router that is also the ancestor of the destination core. It is then routed down to the destination. In this minimum path routing, the upward path is flexible but the downward path is fixed. That is to say, when a packet is routed upward, either left or right output ports can be selected; but when it is routed downward, only one direction is possible to reach the destination. In selection of the upward links, either adaptive or deterministic routing algorithm can be used.

Compared with deterministic routing, adaptive routing owns more choices and thereby with potential of supporting higher throughout. This is, however, not always the case. Adaptive routing would quite likely consume more power to preserve information and perform calculation, and it also takes more time to make decision, which may impair performance. In addition, the packets may be out of order because of selection of different paths, which may ask for a huge buffer at the destination to reorder the data. In addition, reorder time would introduce extra delay. In our fat tree network, links are relative abundant and therefore the advantage of adaptive routing is further reduced. A detailed comparison between adaptive routing and deterministic routing is given in [33].

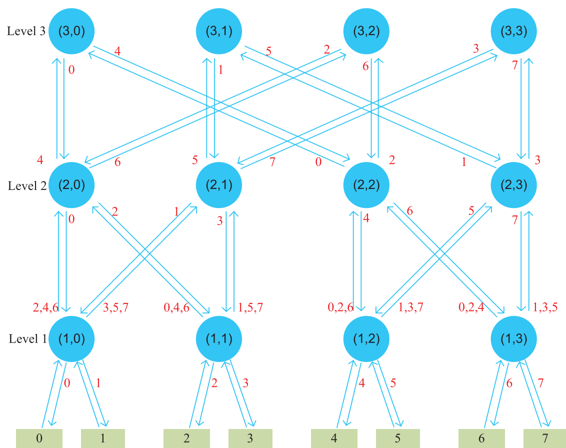


Fig. 6. Deterministic routing in fat tree. The labels attached on the arrows show the allowed destinations of packets.

In our design, the network controller is responsible for all paths configuration, which will be the bottleneck if it cannot process requests fast enough. For the benefit of the overall performance, we should make the network controller simple and fast, and thus deterministic routing is selected in UNION.

In deterministic routing, traffic should be balanced by wise path-selection. It is observed that either ascending link of a router can be chosen in the upward path selection, whereas downward path is always fixed. Therefore, we should distribute upward packets reasonably to avoid collision in the descending links. For example, in Fig. 6, router (1, 3) in level 1 connects two concentrators six and seven and therefore will receive packets destined to them. And, we hope that the packets with different destinations will come from different links, specifically, from the left father router and from the right father router separately. Otherwise there is a collision on the same coming link, leaving another link idle. If the two packets come from different links, we trace back the paths and would always find that these packets choose different ascending links in the level-1 router. This phenomenon was noticed in [34]. If we always let packet destined to 6 (7) select the left (right) ascending link in the level-1 router, it is guaranteed that they would not collide at the router (1, 3). In addition, the arrangement has not increased the unbalance on ascending links as the packets destined to the two cores are evenly distributed in two links.

Based on above observation, we adopt a shuffling technology which is similar to [33]. In our design, unlike packet switching in [33], network controller finds path without probing the path and thus further improve the performance. According to the scheme, in the ascending stage, the packets with neighboring destinations are routed to different output ports. For instance, as in Fig. 6, at router (1, 0), packets with destination 2, 4, 6 will be sent to one upward port whereas packets with destination 1, 3, 5 will be sent to the other upward port. Formally, in router at level i, if the $(i-1)$th bit of the destination is 0, the packet would be routed to the left path, otherwise the right path. In Fig. 6, for a fat tree with eight concentrators, all possible packets on each link are listed. As the path is only determined by the source/destination

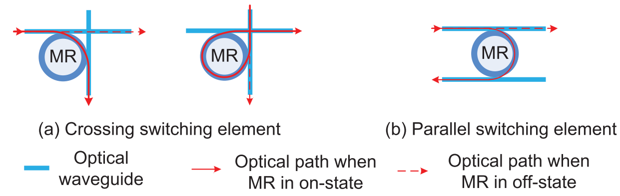


Fig. 7. Two basic switching elements. The direction of optical signal can be controlled by MR.

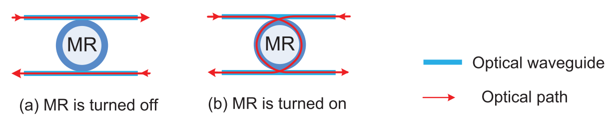


Fig. 8. Single MR working with two optical signals of the same wavelength.

information, the network controller can implement the routing algorithm quickly.

4) Optical Router: Optical routers are responsible for switching optical signals from one port to another port, hence the network can be dynamically configured. They are based on two basic 1×2 switching elements, including the parallel and crossing types. From Fig. 7, both of the switching elements consist of two waveguides and one MR. The resonance wavelength of an MR can be controlled by electrical voltage. When the wavelength of the input light is the same as the resonance wavelength of the MR, resonance happens and the light would be diverted to another waveguide. On the other hand, if resonance does not happen, the light would bypass the MR directly. For both switching elements, two light sources can be added into it simultaneously. From Fig. 8, if resonance does not happens, two light waves would propagate along the original waveguides; if resonance happens, both light waves would be diverted to different waveguides. Therefore, it is actually a 2×2 switching element. This property is discussed in [35], and illustrated with experiment in [36]. MR is wavelength-sensitive and each kind of MR can control corresponding light signals while not affecting the light in other wavelengths. UNION transmits payload data signals and control signals in wavelengths λ_0 and λ_1 separately. Two kinds of signals can be transmitted in the same waveguide without interfering with each other.

Based on the two basic switching elements, we build an optical router, called as the optical turnaround router (OTAR), for the fat tree-based intrachip optical network. Routers are grouped into router clusters, and each cluster as a whole is controlled by an electronic control unit that receives commands from the network controller. All clusters are shown in Fig. 3, and a level-2 cluster consisting of two routers is shown in Fig. 9.

The switching fabric of each OTAR router implements a 4×4 switching function for optical data signals in wavelength λ_0. Based on the routing algorithm, some turns in the router can be omitted. Specifically, there are neither U-turns nor turns between upper left and upper right ports. The routing functions can be achieved by turning on/off corresponding MRs. For example, when the leftmost microresonator is turned on, the light from lower left port would be diverted to upper right

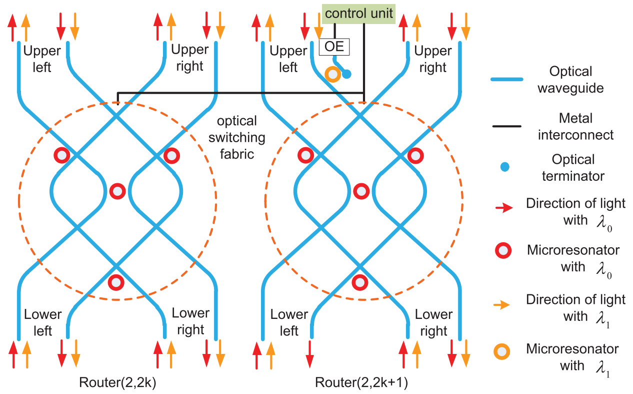


Fig. 9. Level-2 router cluster including two OTARs. Light would propagate along the waveguide unless it is diverted by the MR to another one. All MRs are controlled by an electrical control unit.

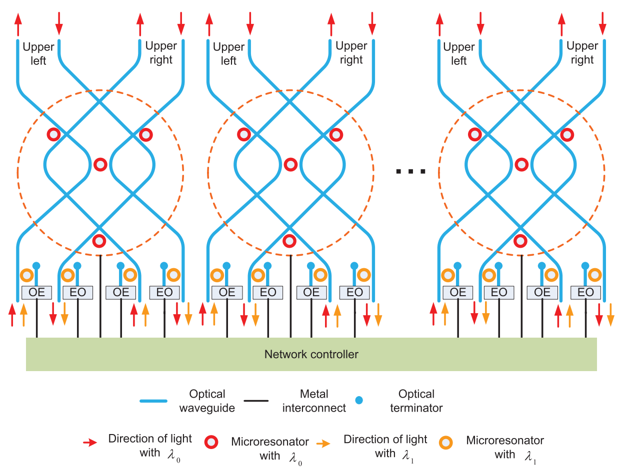


Fig. 10. Top-level routers. The upward ports are to connect interchip networks. Network controller receives request signals from upward waveguides and sends configuration signals to clusters and cores through downward waveguides.

port while the light from lower right port would be diverted to upper left port. When it is powered off, lights would propagate along the original waveguides. Our router is designed to minimize the number of waveguide crossings and MRs.

One of the routers of a cluster is different from the others, and it is with a control signal receiver. These special routers are shown in Fig. 3. From Fig. 9, the right router is the one attached with a control signal receiver. The MR with resonance wavelength λ_1 will direct the control signals from the waveguide to the router control unit. The received control information, sent from the network controller, would be interpreted to configure all MRs in this cluster with resonance wavelength λ_0. After MR configurations, the path is setup for payload data signals in wavelength λ_0.

Top-level routers are shown in Fig. 10. All the top-level routers form a cluster in which the network controller also resides. The switching fabric for data signals is the same as in all other routers as shown in Fig. 9. These routers are, however, attached with MRs with resonance wavelength λ_1, receiving requests originated from processing cores at the leaves of the tree. Control information from the network controller to each cluster and core is sent through an EO interface attached to the downward waveguide as shown in figure.

With the above designs of router and clusters, the network controller is connected to all clusters and concentrators in a point-to-point fashion. Therefore, besides the data network, a control network is also successfully built with the same fabrics. In the other words, a single optical network is used for both data and control information. This saves the network resources and also prevents the potential waveguide crossings between two networks. In the following section, we will show how the interchip network is designed and how it is connected to the intrachip network.

B. Interchip Network

The interchip network connects all intrachip networks. In UNION, we designed an optical bus with distributed control for the interchip interconnections (Fig. 11). Network controllers collaboratively arbitrate the optical bus and manage their own on-chip network resources for interchip communications. Although bus-based communication architecture has limited scalability, it is still a viable low-cost choice for systems with a moderate number of chips. The low-cost design can improve the feasibility of the whole system. Another advantage of using bus is that we need not fix the system size at early stages of design time.

UNION's interchip network consists of an optical data bus (at top of Fig. 11) and an optical control bus (at bottom of Fig. 11). The data bus is responsible for data communications among chips, and the control bus helps network controllers to cooperate with each other during bus arbitration.

1) Optical Data Bus: In UNION's interchip network, the number of data bus channels is proportional to the number of top-level routers in the intrachip network. Specifically, each upward port of the top-level router of fat tree would access a separate bus channel. For 64-core CMPs, only 16 data bus channels are required. Each data bus channel is composed of on-chip silicon waveguides, polymer waveguides embedded on the PCB, and optical connectors connecting on-chip waveguides with on-board waveguides. Each channel is bidirectional and half-duplex. We design interface switches to connect top-level routers in the intrachip network to optical data bus channels, as shown at the top of Fig. 11. The interface switch is composed of MRs and waveguides. Data signals can be sent to the bus in either direction depending upon that MR is powered on. The router can also fetch the data from the bus with the corresponding MR to be powered on. If no MR is powered on, signals will pass current chip with little optical power loss.

To boost the throughput of the bus, we segment the waveguides into multiple sections and each section can support a transaction independently. These sections would not interfere with each other given that MR can divert the light from bus completely. From Fig. 11, for instance, on a specific channel, chip 1 can send data to chip 0 although it is receiving data from chip M. In the best case, a specific channel can be shared by all M chips simultaneously: chip 0 sends data to chip 1, chip 1 sends to chip 2, and so on. This inherent parallelism can improve the network throughput significantly.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

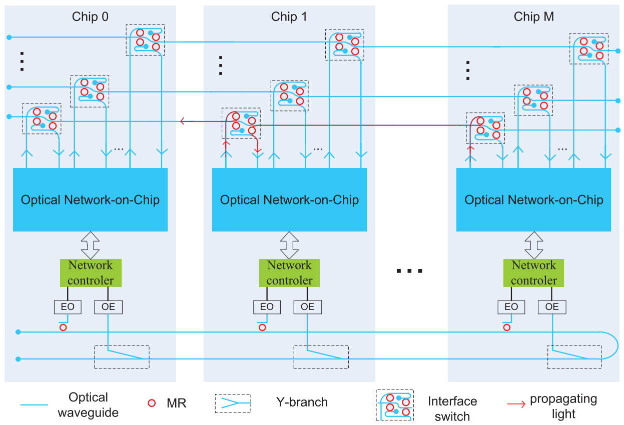


Fig. 11. Interchip optical network. The top-parallel waveguides with interface switches are data bus. Each waveguide is bidirectional for optical signals. The bottom waveguide with Y-branch is a broadcast control bus.

There is no electrical buffering required between on-chip and off-chip interconnects. The E/O conversions would consume a lot of power, and large buffer would occupy extra area and it is also power-hungry. If the buffer size is not large enough, maximum packet size would be restricted. Truncating a large packet into many small packets would introduce large arbitration overhead, impairing the performance of the whole network. Without data buffering between on-chip and off-chip interconnects, the on-chip network need an arbitration scheme to make sure the bus is free to use, which is detailed in next section.

2) Optical Control Bus: As a single channel is shared by multiple chips, arbitration is required to avoid collisions. The bus arbitration is made collaboratively by the network controllers. A control bus is implemented to help them cooperate with each other, shown at the bottom of Fig. 11. The control bus is primarily a waveguide connecting all the network controllers. It allows a network controller to broadcast control signals. From figure, an MR is used to inject control signals into the control bus, and a Y-branch is used to eject control signals. Y-branches are designed with different split ratio. The $(N-i)$th Y-branch along the bus has the split ratio of $i:1$, and this enables each network controller to receive the same amount of power.

A useful property of optical signal is that its propagation delay is predictable. Wave pipelining thus can be implemented here to reduce the delay. Specifically, each chip can send some data out on the same waveguide simultaneously as long as they do not collide with each other. For example, if the distance between two chips is 10 cm, the light propagation speed is c/n where refraction index n equals 1.5, and the data rate is 40 Gb/s, then simultaneously each chip can send 20 bits out. These 20 bits are enough for cooperation communication.

3) Network Protocols: Interchip communications require both intrachip and interchip networks, and are managed collaboratively by the network controllers. When a core wants to start a communication with another core on a different chip, it first sends a request to the network controller through a concentrator, which is the same as an intrachip communication. After receiving the request, the network controller will first try to reserve the upward path from the concentrator to an interface switch on the chip according to the same deterministic routing algorithm as for intrachip communications. After successful reservation of the upward path, the network controller would broadcast the transaction requests by the control bus. After receiving the request, the destination network controller would try to reserve the downward path on the chip from a specific interface switch to destination concentrator. If the downward path is reserved successfully, the destination network controller would try to reserve the interchip bus by broadcasting bus request. All network controllers would receive the bus requests, and they would decide which ones should be granted according to the arbitration scheme we will discuss later. Once the data bus section is reserved, the source network controller would send a grant signal to the source core. Then, the source core will send data out immediately. Upon finishing the data transmission, a tear down signal is sent from the source core to the source network controller, which in turn broadcasts it to the destination controller. All network controllers will update their status buffers based on received information.

We should mention that, as all control signals are broadcasted on the control bus, each network controller would have a complete copy of information of the data buses. Therefore, when bus requests come, each network controller can independently decide whether they should be granted or not based on the exactly same arbitration scheme which will be discussed later. This independent arbitration can save the negotiation delay among network controllers.

4) Bus Requests Arbitration: When multiple bus requests are broadcasted on the control bus channel, network controllers need to make arbitrations efficiently. In our design, each network controller adopts the same arbitration scheme and makes arbitration independently. Round-robin can be used in general bus arbitration. With the unidirectional property of optical signals, multiple requests may share, however, a single bus channel simultaneously. Therefore, we need a better algorithm to take advantage of the property and improve the performance. The algorithm is explained by an example as follows.

From Fig. 12(a), there are eight links labeled by number and eight bus requests labeled by alphabet. The links constitute a bus channel, and each requested transaction would occupy some links. We assume all transactions take the same time. According to the unidirectional property of optical signals, all requests that are not overlapping on the links can be scheduled simultaneously. Our aim is to finish all transactions as soon as possible, and correspondingly in the figure, the aim is to rearrange the wires up or down to minimize the height of wires. We first propose a simple off-line greedy algorithm to schedule the requests. This algorithm is similar to the Interval Partitioning problem, and it is optimal [37]. The algorithm is shown in Algorithm 2. And the scheduling result is shown in Fig. 12(b).

With the above algorithm, we can always find a request group that can be scheduled simultaneously. For online scheduling, new requests may come later, and we have to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Algorithm 2 Bus Arbitration

Require: Requests R

```
i ⇐ 0;
while R ≠ φ do
  for All Requests in R do
    Arbitrary select a request K in R
    if K ⋂ Group[i] == φ then
      Group[i] ⇐ Group[i] ⋃ K
      R ⇐ R − K
    end if
  end for
  i++;
end while
return Group
```

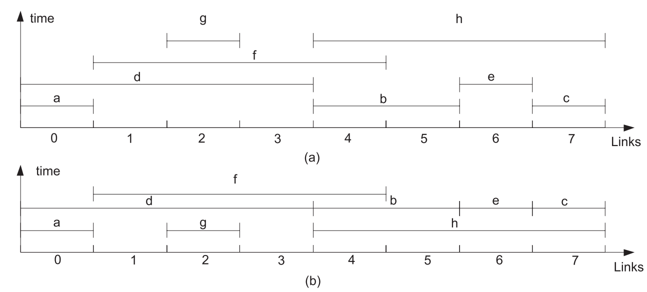


Fig. 12. Bus arbitration algorithm example. The scheduling result (b) shows that proposed scheme finishes all transactions faster than original one (a).

run the algorithm again. This is not optimal any more. It is, however, in fact near optimal if there is few requests coming before old ones are scheduled out. Because of its simplicity and relative effectiveness, we adopt this algorithm in our design.

5) Adaptive Power Control Mechanism: We propose an adaptive power control mechanism to improve the power efficiency of the network, and it is applicable to other optical NoC architectures as well. The mechanism is based on the observation that a large portion of power is consumed by lasers in the network. For example, in an 80-nm design, while the power consumed by a simple data link to transfer one bit is about 2.5 pJ, the laser source consumes about 1.68 pJ which considers a large proportion [38]. Our control scheme is to reduce the power consumption of laser dynamically.

To transfer the data over a link successfully, the emission power from the laser should be larger than the summation of power loss along the path and minimum optical power required at the destination for sufficient large SNR. Traditionally, to guarantee enough power for all possible transmissions, the worst case power loss in the optical interconnects is considered, and laser sources are set to provide the worst case optical power for all packets. This also causes the destination circuits to receive optical power within a large dynamic range. In addition, too much power on one transaction may introduce large noise power leaked into other transactions.

The adaptive power control mechanism we implement here uses routing information to calculate the optical power loss encountered on an optical path and control laser source to generate just-enough optical power for transmission. As the routing path is already decided by the network controller before data transmission, the optical power loss can be easily obtained. As the deterministic routing algorithm is adopted here, a precalculated table can be used. While the network controller is trying to setup an optical path, the concentrator would calculate the minimum launch power of the laser and drive the VCSEL with appropriate driving current. The turn-on delay of VCSEL is below 1 ns [39] and it can overlap with path setup delay that takes more than 10 ns. After VCSEL is biased above threshold, the direct modulation speed is 40 Gb/s [40], [41]. Compared with nonadaptive mechanisms, the adaptive power control mechanism avoids unnecessary power consumption and improves the power efficiency of UNION.

IV. Evaluation and Results

We compare UNION with its matched electrical counterpart for performance, energy consumption, and delay. UNION is a unified inter/intrachip network, and thus we compare it with the counterpart with both interchip and intrachip networks. The target system is with eight chips, and each chip with 64 cores. The technology is targeted at 45 nm, and the frequency of electrical components including routers and network controller is 1.25 GHz. The chip size is assumed to be 1 cm $\times$ 1 cm. For both architectures, fat tree topology is used for on-chip networks, and bus is for interchip interconnects. In UNION, the concentrators are plugged into the four-level high-optical fat tree; in electrical counterpart, the cores are directly plugged into the six-level high-electrical fat tree. The same turn around routing algorithm is used in the on-chip networks. The bandwidth of on-chip links in both networks is assumed to be the same 40 Gb/s. We also assume the same bisectional bandwidth of interchip buses.

In the electrical fat tree NoC, packet switching is used. Wormhole routing is adopted to reduce the packet delay. The electronic routers are pipelined with three cycles delay. Two virtual channels are implemented to avoid the head-of-line problem and improve performance. Back pressure is used for flow control. Each port of the router is 32-bit wide and bidirectional. Link delay is confined in one cycle, and thus the link bandwidth is 40 Gb/s. The top-level routers of electrical fat tree are also connected to electrical interchip buses with the topology similar to UNION. But as there are bandwidth gap between on-chip and interchip link, serializers/deserializers are required at the interfaces. We assume each bus channel works at 10 Gb/s [42]. There are 64 bidirectional channels connecting 32 top-level routers, and thus the bisectional bandwidth of the bus is 640 Gb/s. It is possible to use other technologies to build interchip network, e.g., 3-D stacking. 3-D stacking is an attractive approach to connecting chips with low-latency TSVs, and it would be discussed in our future work.

For comparison, we assume the link bandwidth in UNION is also 40 Gb/s. As the 16 concentrators, instead of 64 cores, are connected with the intrachip optical network, the bisectional bandwidth of optical NoC is only a quarter of the electrical NoC. There are 16 bidirectional data bus channels, providing

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

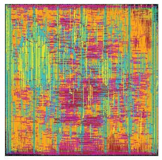


Fig. 13. Layout of network controller. It runs at 1.25 GHz, consuming 30.6 mW with switching rate of 20%. The area is 112951.16 μm^2.

640 Gb/s bisectional bandwidth which is the same as the electrical bus. We assume transmission delay of silicon waveguide is 140 ps/cm [43]. The interlayer delay between stacked chips was 16 ps as modeled in [44], and the power consumption of transmitting each bit is 0.02 pJ/bit [45]. We implement the network controller in VHDL and synthesized it with a 45-nm library. Synthesis results show that the network controller can simultaneously process 16 requests in 20 clock cycles. This implies that the network controller can be applied in a much larger system. For example, in a 256-core system, the 64 concentrators with injection rate 0.3 would generate 0.6 request/cycle, given the packet size is 32 flits. This request rate can be tackled by the same network controller. To support even larger system, we can increase the parallelism level in the network controller, and it would be further studied in our future work. The layout of the network controller is as shown in Fig. 13.

We use detailed cycle-accurate simulators programmed in SystemC to study the performance of both architectures. The power model is also embedded in the simulators to evaluate the power efficiency of the networks. The simulations are based on a set of real-CMP applications, including H264 decoder with different rates, satellite receiver, sample rate converter, fpppp, sparse matrix solver, and robot control. Fixed access pattern for each application is studied, and an off-line optimization approach is applied for mapping and scheduling tasks onto the CMP with the objective of maximizing system performance [46]. We assign the tasks to the cores and minimize the total amount communication volume. Communication locality is maximized to reduce network congestions caused by the interferences among different transactions.

A. Performance Comparison

Performance is measured in terms of the average number of iterations that an application can be finished in a given time. Fig. 14 shows the performances of each application on CMPs using UNION compared with the electrical counterpart. For most applications, CMPs using UNION achieve more than 3× improvement compared with the CMPs using its electrical counterpart. The satellite receiver application is an exception that only 10% improvement is achieved by UNION. This is because that under this application, most traffics are on-chip transmissions. When most of the data flows are confined on an individual chip, the contribution of the unified design would not be well illustrated.

The satellite receiver application, with most of traffics on chip, also shows the efficiency of our central-control protocol, considering that the bisection bandwidth of UNION on chip

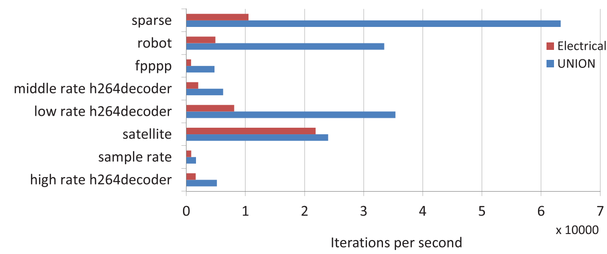


Fig. 14. Performances of UNION and electrical counterpart under real applications. The performances are shown in terms of execution iterations/s.

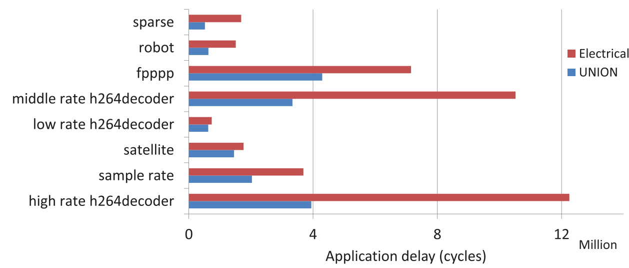


Fig. 15. Application delays of real applications in the UNION and the electrical counterpart.

is only a quarter of electrical counterpart. With central-control protocol, the path setup delay is small because of the low latency of light transmission. More importantly, if collision happens during path setup in network controller, the unscheduled requests would not occupy any network resources. On the other hand, in the referenced electrical network, the blocked packets may in turn block other packets. WDM technology may also be used in UNION to potentially improve the bandwidth, and it will be considered in our further work.

Fig. 15 shows the application delays of each traffic in UNION compared with the matched electrical network. The application delay is the time between the start and completion of an application. On average, the execution time of UNION is 52% of its electrical counterpart.

B. Energy Evaluation and Comparison

We evaluate the energy efficiency of UNION and its matched electrical network. The energy efficiency is measured as the average energy consumption for transferring per bit in the network. All the electrical devices are target at 45-nm process. In UNION, for a transmission within the same concentrator, energy consumption includes the energy required to transfer the packet through the two local electrical interconnects, the energy dissipated by the local electrical switching fabric and the energy consumed by the control unit. For a transmission between concentrators, besides the energy consumed by local concentrators, additional energy is required for optical transmission of control signals and payload data.

The necessary optical power emission of the laser is estimated as the summation of optical power loss in the path and the minimum optical power required at the destination. Therefore, in addition to improving the device technologies of optical transceivers, O/E power efficiency can also be improved by reducing the optical power loss encountered in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

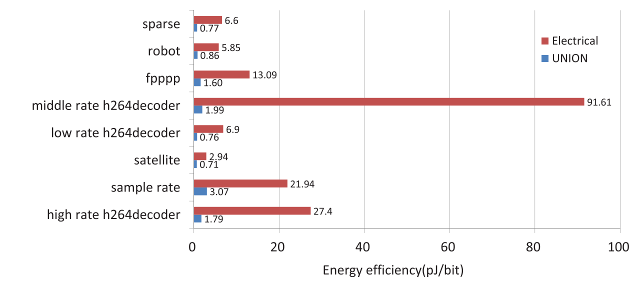


Fig. 16. Energy efficiency of UNION and electrical counterpart under real applications.

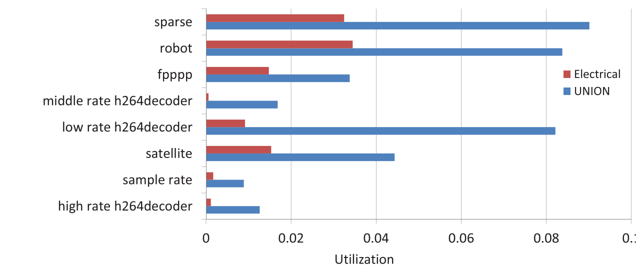


Fig. 17. Aggregate switching capacity utilization of UNION and electrical counterpart under real applications.

the optical link, which is one of the reasons why we reroute the waveguides to avoid crossings in the floorplan. The optical loss of each component is assumed as below. The silicon waveguide crossing loss, MR insertion loss, MR passing loss, waveguide bending loss, and waveguide propagation loss are 0.12, 0.5, 0.005 dB, 0.005 dB/90° and 0.17 dB/mm, respectively, [47]–[49]. The coupling loss between on-chip and on-board waveguides is 0.45 dB [50]. The propagation loss on the polymer waveguide on PCB is 0.035 dB/cm [51].

The power models of the driver, TIA-LA circuit, serializer, and deserializer are derived from [38], [52]. The VCSEL model is derived from [53]. As we are comparing with 45-nm electrical circuits, all the related power consumptions are linearly scaled to 45 nm. Specifically, the driver and TIA-LA circuits power consumption is 0.46 pJ/bit, and the power consumption of serializer and deserializer is 0.288 pJ/bit. For the photodetector, we assume the sensitivity is -14.2 dBm with bit error rate of 10^{-12} [54]. We also assumed 1 μW/K/ring heating power, and 20-K tuning range as in [29]. The pitch width of waveguides on PCB is assumed to be 25 μm, whereas the pitch width of waveguides on chip is assumed to be 5.5 μm [5]. The spacing is large enough to avoid coupling loss between waveguides.

As for the electrical network, the electronic router and metal wires are simulated in Cadence Spectre, and power characteristics are derived based on the simulations. Simulation results show that on average the crossbar consumes 0.06 pJ/bit, the input buffer consumes 0.003 pJ/bit, and the control unit consumes 1.5 pJ to make decisions for each packet. For the off-chip electrical wires, low-swing signaling technology is used and we used the latest power consumption results from [42].

Fig. 16 shows the energy consumption of UNION compared with the electrical counterpart for different applications. On average, UNION consumes 88% less energy compared with the matched electrical network. Further analysis shows that, long metallic interconnects and buffers consume a large amount of power in electrical network, while all these are omitted in optical network. The seamless connecting between on-chip and interchip optical link makes UNION consume very little power consumption for interchip communication compared with the electrical counterpart. The adaptive power control mechanism further improves UNION's energy efficiency. Besides the optical interconnects, UNION also benefits from the short electrical wires in power consumption. Lots of localized intraconcentrator traffics in UNION help to bring the average power consumption even lower than 1 pJ/bit. It is shown in figure that for the satellite receiver application, both optical and electrical networks achieve the lowest power consumption. This is in accordance with the fact that most traffics of the application are intrachip communications.

C. Network Resource Analysis

Table I shows the resources we allocate to each chip in UNION. The area of a single optical router is about 1600 μm^2, with 12-μm-diameter MRs. For a CMP chip size of 10 $\times$ 10 mm^2, the total area of waveguide and optical switching fabrics is about 0.45 mm^2.

To show the capability of the network, aggregate switching capacity is studied here. The switching capacity shows how much bandwidth can be switched by a component such as router. For example, a 4 $\times$ 4 electronic router with each port bandwidth 40 Gb/s, the switching capacity of this router is 160 Gb/s as it can switch four ports simultaneously at best. The aggregate switching capacity of the whole network is the summation of switching capacity of all switching components. A network with higher aggregate switching capacity indicates that it consumes more resources and intends to support higher throughput in theory. The utilization of aggregate switching capacity can help to clarify the efficiency of the network. For example, if there is no data passing through a specific router, then the switching capacity of this router is wasted. In our simulation, we try to detect this parameter to evaluate the efficiency of the network. The result is shown in Fig. 17. As we can see, UNION utilizes network resources much more efficiently than the electrical network for most applications.

To show the design tradeoffs for on-chip network, we compare UNION with two alternative architectures including Corona [6] and optical Clos network [23]. UNION is a unified design with both on-chip and off-chip networks, whereas both Corona and Clos are only on-chip network designs. In the following comparison, we only consider the on-chip portion of UNION that is a fat tree network with centralized control mechanism. Corona is an optical crossbar targeting a CMP with 256 cores. The optical Clos network is designed for a 64-tile system. For comparison with UNION, we scale the Corona and optical Clos architectures to a CMP with 64 cores, and each channel is with bandwidth of 40 Gb/s. Every 4 cores are concentrated as a cluster and the communication within the cluster is omitted such that we can focus on the design of optical network. The resulted networks are a 16×16 Coronalike crossbar and a four-ary, three-stage Clos network. The power models of the optical and electrical circuits are

TABLE I
COMPARISON OF UNION, CORONALIKE CROSSBAR [6], AND OPTICAL CLOS [23]

	UNION	Crossbar	Clos
4 × 4 Optical router	32	0	0
4 × 4 Electrical router	0	0	12
5 × 5 Electrical router	16	16	16
Laser source	46	1 (off-chip)	2 (off-chip)
Photodetector	46	320	96
MR	142	1280	192
Maximum throughput (uniform)	180 Gb/s	640 Gb/s	309 Gb/s
Energy efficiency	0.94 pJ/bit	3.58 pJ/bit	3.52 pJ/bit

assumed the same as in UNION. Besides multistage networks like Clos and one-stage crossbar, there are other possible topologies such as small-world network [55]. Comparing with the conventional planar network, the small-world network reduces the diameter of the network and thus, the transmission stages for the traffic. Implementing an optical small-world network for on-chip network and incorporating it into the off-chip network for a unified design would be further explored in our future work.

The comparison results are shown in Table I. Because of resource contention, UNION can support around 28% throughput achieved by the crossbar. The high throughput of the crossbar is at the cost of high network resources requirement from Table I. A fully connected crossbar would also make the optical path more complex such that larger optical loss is encountered for the light. Therefore, higher optical power is consumed. Under the uniform traffic with injection rate of 0.2, the energy efficiency of UNION is around 0.94 pJ/bit versus 3.58 pJ/bit of the crossbar. Clos network uses electrical routers for packet switching and optical links connecting routers. It can support 48% throughput of the crossbar. Comparing with fully connected crossbar, Clos reduces the complexity of the optical path, and hence, the optical power. The costs are the electrical switching elements and relative long electrical wires connecting clusters to routers, which are power consuming. The power efficiency of Clos is 3.52 pJ/bit that is larger than UNION but smaller than crossbar. In conclusion, compared with crossbar and Clos, UNION saves 74% and 73% of energy for transferring a bit, respectively. It also achieves 28% of the crossbar's throughput with only 11% of MRs.

D. Scalability Analysis

Here, we compare UNION with the matched electrical network for scalability. To test the scalability, we gradually increase the number of chips in the targeted system. Each chip is the same with 64 cores. A system with heterogeneous chips would be considered in our future work.

We simulate both networks for the zero-load latency. The packet size is 1024 flits, and each flit is 32 bits. The energy efficiency is measured as the average energy consumption for transferring per bit in the network using all possible links. The packet delay comparison between UNION and electrical network is shown in Fig. 18, and the energy comparison

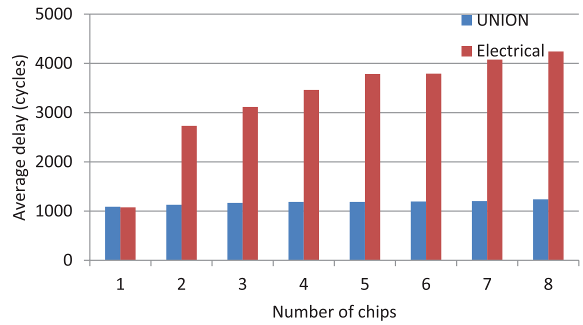


Fig. 18. Zero-load latency comparison between UNION and matched electrical network.

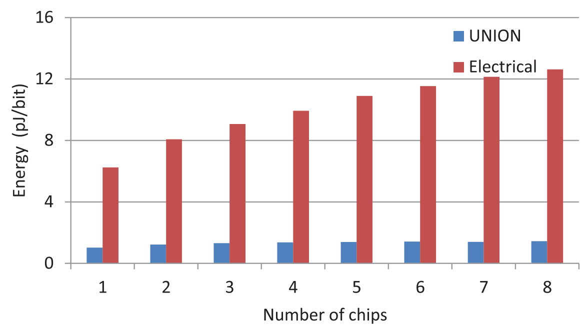


Fig. 19. Average energy consumption of UNION and matched electrical network.

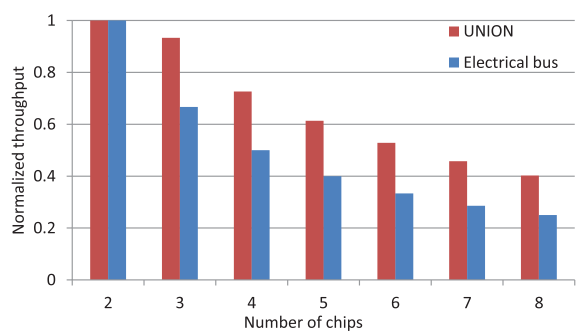


Fig. 20. Normalized available throughput for each chip in the interchip networks of UNION and electrical counterpart.

is shown in Fig. 19. In UNION, the average packet delay and energy consumption increase very slightly respecting to the number of chips, showing very good scalability. But in the matched electrical network, the packet delay and energy consumption increase quickly. The performance gap between UNION and the electrical network is widened with larger number of chips. In addition, for the electrical network, there is a giant leap between the one- and two-chip system, showing that there is a huge performance gap between on-chip and off-chip interconnects. This phenomenon does not exist in UNION, and the reasons are as follows. For optical network, interchip propagation delay is very low and the arbitration delay is independent of the hops. For electrical network, interchip traversing involves much larger propagation delay and the serialization delay. Similarly, the power loss of the optical signal on the interchip links is not significant, while the power consumption of the electrical IO is much higher compared with on-chip wires.

For the bus interconnect, the available throughput for each chip would drop as the number of connected chips increases. In the electrical bus, the available throughput per chip is the inverse of the number of chips, given that the bus can only be used by one transaction at a time. In contrast, the data bus of UNION is divided into multiple independent sections and they can support multiple transactions simultaneously. Therefore, the available throughput for each chip drops slower than the electrical counterpart, as shown in Fig. 20. The limit of the number of chips is decided by the acceptable throughput required by the applications, and our data channel design improves the scalability.

V. Conclusion

In this paper, a unified inter/intrachip optical interconnection network, called UNION, was proposed for CMPs. It employed a hierarchical optical network to separate interchip communication traffic from intrachip communication traffic. It fully utilized a single optical network to transmit both payload and control packets. The network controller on each CMP not only managed intrachip communications, but also collaborated with each other to facilitate interchip communications. We compared CMPs using UNION with those using a matched electrical counterpart in 45-nm process. Simulation results of eight applications showed that on average UNION improved CMP performance by 3× while reducing 88% of network energy consumption.

References

[1] W. J. Dally and B. Towles, "Route packets, not wires: On-chip interconnection networks," in *Proc. Design Autom. Conf.*, 2001, pp. 684–689.

[2] J. Xu, W. Wolf, J. Henkel, and S. Chakradhar, "H.264 HDTV decoder using application-specific networks-on-chip," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2005, pp. 1508–1511.

[3] F. E. Doany, C. L. Schow, R. Budd, C. Baks, D. M. Kuchta, P. Pepeljugoski, J. A. Kash, F. Libsch, R. Dangel, F. Horst, and B. J. Offrein, "Chip-to-chip board-level optical data buses," in *Proc. Nat. Fiber Opt. Eng. Conf., Opt. Fiber Commun.*, 2008, pp. 1–3.

[4] I. A. Young, E. Mohammed, J. T. S. Liao, A. M. Kern, S. Palermo, B. A. Block, M. R. Reshotko, and P. L. D. Chang, "Optical I/O technology for tera-scale computing," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 235–248, Jan. 2010.

[5] N. Kirman, M. Kirman, R. K. Dokania, J. F. Martinez, A. B. Apsel, M. A. Watkins, and D. H. Albonesi, "Leveraging optical technology in future bus-based chip multiprocessors," in *Proc. 39th Annu. IEEE/ACM Int. Symp. Microarchit.*, Dec. 2006, pp. 492–503.

[6] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn, "Corona: System implications of emerging nanophotonic technology," in *Proc. 35th Int. Symp. Comput. Archit.*, 2008, pp. 153–164.

[7] A. Shacham, K. Bergman, and L. P. Carloni, "Photonic networks-on-chip for future generations of chip multiprocessors," *IEEE Trans. Comput.*, vol. 57, no. 9, pp. 1246–1260, Sep. 2008.

[8] M. J. Cianchetti, J. C. Kerekes, and D. H. Albonesi, "Phastlane: A rapid transit optical routing network," in *Proc. 36th Annu. Int. Symp. Comput. Archit.*, 2009, pp. 441–450.

[9] J. M. Perkins, T. L. Simpkins, C. Warde, and J. C. G. Fonstad, "Full recess integration of small diameter low threshold VCSELs within Si-CMOS ICs," *Opt. Exp.*, vol. 16, no. 18, pp. 13955–13960, 2008.

[10] A. V. Krishnamoorthy, K. W. Goossen, W. Jan, X. Zheng, R. Ho, G. Li, R. Rozier, F. Liu, D. Patil, J. Lexau, H. Schwetman, D. Feng, M. Asghari, T. Pinguet, and J. E. Cunningham, "Progress in low-power switched optical interconnects," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 357–376, Mar.–Apr. 2011.

[11] G. Van Steenberge, P. Geerinck, S. Van Put, J. Van Koetsem, H. Ottevaere, D. Morlion, H. Thienpont, and P. Van Daele, "MT-compatible laser-ablated interconnections for optical printed circuit boards," *J. Lightw. Technol.*, vol. 22, no. 9, pp. 2083–2090, Sep. 2004.

[12] R. Yoshimura, M. Hikita, M. Usui, S. Tomaru, and S. Imamura, "Polymeric optical waveguide films with 45° mirrors formed with a 90° V-shaped diamond blade," *Electron. Lett.*, vol. 33, no. 15, pp. 1311–1312, Jul. 1997.

[13] S. H. Hwang, M. H. Cho, S.-K. Kang, H.-H. Park, H. S. Cho, S.-H. Kim, K.-U. Shin, and S.-W. Ha, "Passively assembled optical interconnection system based on an optical printed-circuit board," *IEEE Photon. Technol. Lett.*, vol. 18, no. 5, pp. 652–654, Mar. 2006.

[14] A. Apsel, Z. Fu, and A. G. Andreou, "A 2.5-mW SOS CMOS optical receiver for chip-to-chip interconnect," *J. Lightw. Technol.*, vol. 22, no. 9, pp. 2149–2157, Sep. 2004.

[15] M. P. Christensen, P. Milojkovic, M. J. McFadden, and M. W. Haney, "Multiscale optical design for global chip-to-chip optical interconnections and misalignment tolerant packaging," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, no. 2, pp. 548–556, Mar.–Apr. 2003.

[16] C. Berger, M. Kossel, C. Menolfi, T. Morf, T. Toifl, and M. Schmatz, "High-density optical interconnects within large-scale systems," *Proc. SPIE*, vol. 4942, pp. 222–235, Apr. 2003.

[17] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. Holzwarth, M. Popovic, H. Li, H. Smith, J. Hoyt, F. Kartner, R. Ram, V. Stojanovic, and K. Asanovic, "Building manycore processor-to-DRAM networks with monolithic silicon photonics," in *Proc. 16th IEEE Symp. High Perform. Interconnects*, Aug. 2008, pp. 21–30.

[18] P. Koka, M. O. McCracken, H. Schwetman, X. Zheng, R. Ho, and A. V. Krishnamoorthy, "Silicon-photonic network architectures for scalable, power-efficient multi-chip systems," in *Proc. 37th Annu. Int. Symp. Comput. Archit.*, 2010, pp. 117–128.

[19] P. Kapur and K. C. Saraswat, "Comparisons between electrical and optical interconnects for on-chip signaling," in *Proc. IEEE Int. Interconnect Technol. Conf.*, 2002, pp. 89–91.

[20] R. G. Beausoleil, P. J. Kuekes, G. S. Snider, S.-Y. Wang, and R. S. Williams, "Nanoelectronic and nanophotonic interconnect," *Proc. IEEE*, vol. 96, no. 2, pp. 230–247, Feb. 2008.

[21] S. Pasricha and N. Dutt, "ORB: An on-chip optical ring bus communication architecture for multi-processor systems-on-chip," in *Proc. Asia South Pacific Design Autom. Conf.*, 2008, pp. 789–794.

[22] I. O'Connor, "Optical solutions for system-level interconnect," in *Proc. Int. Workshop Syst. Level Interconnect Predict.*, 2004, pp. 79–88.

[23] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic, "Silicon-photonic clos networks for global on-chip communication," in *Proc. 3rd ACM/IEEE Int. Symp. Netw.-Chip*, May 2009, pp. 124–133.

[24] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary, "Firefly: Illuminating future network-on-chip with nanophotonics," in *Proc. Int. Symp. Comput. Archit.*, 2009, pp. 429–440.

[25] Y. Pan, J. Kim, and G. Memik, "FlexiShare: Channel sharing for an energy-efficient nanophotonic crossbar," in *Proc. IEEE 16th Int. Symp. High Perform. Comput. Archit.*, Jan. 2010, pp. 1–12.

[26] X. Wu, Y. Ye, W. Zhang, W. Liu, M. Nikdast, X. Wang, and J. Xu, "UNION: A unified inter/intra-chip optical network for chip multi-processors," in *Proc. IEEE/ACM Int. Symp. Nanoscale Archit.*, Jun. 2010, pp. 35–40.

[27] C. E. Leiserson, Z. S. Abuhamdeh, D. C. Douglas, C. R. Feynman, M. N. Ganmukhi, J. V. Hill, D. Hillis, B. C. Kuszmaul, M. A. St. Pierre, D. S. Wells, M. C. Wong, S.-W. Yang, and R. Zak, "The network architecture of the Connection Machine CM-5," in *Proc. 4th Annu. ACM Symp. Parallel Algorithms Archit.*, 1992, pp. 272–285.

[28] A. Komornicki G. Mullen-Schulz and D. Landon, "Roadrunner: Hardware and software overview," IBM, New York, NY, USA, Tech. Rep., 2009.

[29] P. Guerrier and A. Greiner, "A generic architecture for on-chip packet-switched interconnections," in *Proc. Design, Autom. Test Eur. Conf. Exhibit.*, 2000, pp. 250–256.

[30] P. P. Pande, C. Grecu, A. Ivanov, and R. Saleh, "Design of a switch for network on chip applications," in *Proc. Int. Symp. Circuits Syst.*, vol. 5. 2003, pp. V-217–V-220.

[31] Y. Xie, M. Nikdast, J. Xu, W. Zhang, Q. Li, X. Wu, Y. Ye, X. Wang, and W. Liu, "Crosstalk noise and bit error rate analysis for optical network-on-chip," in *Proc. 47th ACM/IEEE Design Autom. Conf.*, Jun. 2010, pp. 657–660.

[32] H. Gu, J. Xu, and W. Zhang, "A low-power fat tree-based optical network-on-chip for multiprocessor system-on-chip," in *Proc. Design, Autom. Test Europe Conf. Exhibit.*, 2009, pp. 3–8.

[33] C. Gomez, F. Gilabert, M. E. Gomez, P. Lopez, and J. Duato, "Deterministic versus adaptive routing in fat-trees," in *Proc. Int. Parallel Distrib. Process. Symp.*, 2007, p. 292.

[34] Z. Ding, R. R. Hoare, A. K. Jones, and R. G. Melhem, "Level-wise scheduling algorithm for fat tree interconnection networks," in *Proc. ACM/IEEE SC Conf.*, Nov. 2006, p. 9.

[35] H. Simos, C. Mesaritakis, D. Alexandropoulos, and D. Syvridis, "Dynamic analysis of crosstalk performance in microring-based add/drop filters," *J. Lightw. Technol.*, vol. 27, no. 12, pp. 2027–2034, Jun. 2009.

[36] M. Geng, L. Jia, L. Zhang, L. Yang, P. Chen, T. Wang, and Y. Liu, "Four-channel reconfigurable optical add-drop multiplexer based on photonic wire waveguide," *Opt. Exp.*, vol. 17, no. 7, pp. 5502–5516, 2009.

[37] J. Kleinberg and E. Tardos, *Algorithm Design*. Reading, MA, USA: Addison-Wesley, 2005.

[38] C. Kromer, G. Sialm, C. Berger, T. Morf, M. L. Schmatz, F. Ellinger, D. Erni, G.-L. Bona, and H. Jackel, "A 100-mW 4×10 Gb/s transceiver in 80-nm CMOS for high-density optical interconnects," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2667–2679, Dec. 2005.

[39] M. Bruensteiner and G. C. Papen, "Extraction of VCSEL rate-equation parameters for low-bias system simulation," *IEEE J. Sel. Topics Quantum Electron.*, vol. 5, no. 3, pp. 487–494, May–Jun. 1999.

[40] T. Anan, N. Suzuki, K. Yashiki, K. Fukatsu, H. Hatakeyama, T. Akagawa, K. Tokutome, and M. Tsuji, "High-speed 1.1-μm-range InGaAs VCSELs," in *Proc. Opt. Fiber Commun., Nat. Fiber Opt. Eng. Conf.*, 2008, pp. 1–3.

[41] J. A. Lott, N. N. Ledentsov, V. A. Shchukin, A. Mutig, S. A. Blokhin, A. M. Nadtochiy, G. Fiol, and D. Bimberg, "850 nm VCSELs for up to 40 Gbit/s short reach data links," in *Proc. Conf. Lasers Electro-Opt. Quantum Electron. Laser Sci. Conf.*, 2010, pp. 1–2.

[42] G. Balamurugan, J. Kennedy, G. Banerjee, J. E. Jaussi, M. Mansuri, F. O'Mahony, B. Casper, and R. Mooney, "A scalable 5-15 Gbps, 14–75 mW low-power I/O transceiver in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 43, no. 4, pp. 1010–1019, May 2008.

[43] E. Dulkeith, F. Xia, L. Schares, W. M. J. Green, and Y. A. Vlasov, "Group index and group velocity dispersion in silicon-on-insulator photonic wires," *Opt. Exp.*, vol. 14, no. 9, pp. 3853–3863, 2006.

[44] B. S. Feero and P. P. Pande, "Networks-on-chip in a three-dimensional environment: A performance evaluation," *IEEE Trans. Comput.*, vol. 58, no. 1, pp. 32–45, Jan. 2009.

[45] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, "TSV-aware interconnect length and power prediction for 3D stacked ICs," in *Proc. IEEE Int. Interconnect Technol. Conf.*, Jun. 2009, pp. 26–28.

[46] W. Liu, M. Yuan, X. He, Z. Gu, and X. Liu, "Efficient SAT-based mapping and scheduling of homogeneous synchronous dataflow graphs for throughput optimization," in *Proc. Real-Time Syst. Symp.*, 2008, pp. 492–504.

[47] A. W. Poon, F. Xu, and X. Luo, "Cascaded active silicon microresonator array cross-connect circuits for WDM networks-on-chip," *Proc. SPIE*, vol. 6898, p. 689812, Feb. 2008.

[48] S. Xiao, M. H. Khan, H. Shen, and M. Qi, "Multiple-channel silicon micro-resonator based filters for WDM applications," *Opt. Exp.*, vol. 15, no. 12, pp. 7489–7498, Jul. 2007.

[49] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," *Nature Photon.*, vol. 1, no. 1, pp. 65–71, 2007.

[50] J. K. Doylend and A. P. Knights, "Design and simulation of an integrated fiber-to-chip coupler for silicon-on-insulator waveguides," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, no. 6, pp. 1363–1370, Dec. 2006.

[51] G. L. Bona, B. J. Offrein, U. Bapst, C. Berger, R. Beyeler, R. Budd, R. Dangel, L. Dellmann, and F. Horst, "Characterization of parallel optical-interconnect waveguides integrated on a printed circuit board," *Proc. SPIE*, vol. 5453, pp. 134–141, Sep. 2004.

[52] J. Poulton, R. Palmer, A. M. Fuller, T. Greer, J. Eyles, W. J. Dally, and M. Horowitz, "A 14-mW 6.25Gbps Transceiver in 90-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 12, pp. 2745–2757, Dec. 2007.

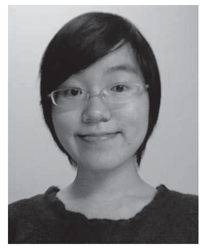
[53] A. Syrbu, A. Mereuta, V. Iakovlev, A. Caliman, P. Royo, and E. Kapon, "10 Gbps VCSELs with high single mode output in 1310 nm and 1550 nm wavelength bands," in *Proc. Opt. Fiber Commun., Nat. Fiber Opt. Eng. Conf.*, 2008, pp. 1–3.

[54] G. Masini, G. Capellini, J. Witzens, and C. Gunn, "A 1550 nm, 10 Gbps monolithic optical receiver in 130 nm CMOS with integrated Ge waveguide photodetector," in *Proc. 4th IEEE Int. Conf. Group IV Photon.*, Sep. 2007, pp. 1–3.

[55] A. Ganguly, K. Chang, S. Deb, P. P. Pande, B. Belzer, and C. Teuscher, "Scalable hybrid wireless network-on-chip architectures for multicore systems," *IEEE Trans. Comput.*, vol. 60, no. 10, pp. 1485–1502, Oct. 2011.

Xiaowen Wu (S'12) received the B.Sc. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 2008. He is currently pursuing the Ph.D. degree in electronic and computer engineering with the Hong Kong University of Science and Technology, Hong Kong.

His current research interests include embedded systems, multiprocessor systems, and network-onchip.



Yaoyao Ye (S'09) received the B.S. degree in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2008. She is currently pursuing the Ph.D. degree in electronic and computer engineering with the Hong Kong University of Science and Technology, Hong Kong.

Her current research interests include network-on-chip, multiprocessor system-onchip, and embedded system.

Jiang Xu (S'02–M'07) received the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 2007.

He was a Research Associate with Bell Labs, Murray Hill, NJ, USA, from 2001 to 2002. He was a Research Associate with NEC Laboratories America, Inc., Cupertino, CA, USA, from 2003 to 2005. He joined a startup company, Sandbridge Technologies, NY, USA, from 2005 to 2007 and developed and implemented two generations of NoC-based ultra-low power multiprocessor systems-on-chip for mobile platforms. In 2007, he joined the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, as an Assistant Professor, and he established the Mobile Computing System Lab. He has authored or co-authored more than 60 book chapters and papers in peer reviewed journals and international conferences. His current research interests include network-on-chip, multiprocessor system-on-chip, embedded system, computer architecture, low-power VLSI design, and HW/SW co-design.

Dr. Xu serves as an Associate Editor of the ACM TRANSACTIONS ON EMBEDDED COMPUTING SYSTEMS and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS. He is an ACM Distinguished Speaker and a Distinguished Visitor of the IEEE Computer Society. He served on the organizing committees and technical program committees of many international conferences.

Wei Zhang, photograph and biography are not available at the time of publication.

Weichen Liu, photograph and biography are not available at the time of publication.

Mahdi Nikdast, photograph and biography are not available at the time of publication.

Xuan Wang, photograph and biography are not available at the time of publication.