Abstract

Photonic Network-on-Chips (NoCs) have recently been proposed due to their inherent low latency and high bandwidth. However, the high static power of the photonic components (e.g., laser source, resonators and waveguides) often results in energy-inefficient architectures. In this paper, we advocate the Energy-Star Photonic Network (ESPN) architecture that optimizes energy utilization via a two-pronged approach: (1) by enabling dynamic resource provisioning, ESPN adapts photonic network resources based on runtime traffic characteristics and (2) by utilizing all-optical adaptive routing. ESPN improves energy efficiency by intelligently exploiting existing network resources without introducing high latency and power hungry auxiliary routing mechanisms. Our evaluation results show that compared to the baseline design, ESPN reduces power and energy consumption under synthetic traffic patterns by 50% and 58% respectively.

Keywords
Photonic, network-on-chip, energy efficient

1. Introduction

With the emergence of multi- and many-core processors, the required bandwidth to support effective on-chip communication is expected to grow rapidly. According to ITRS [1], conventional electrical interconnect will become a power and performance bottleneck for future on-chip communication. As a result, photonic NoCs are drawing increased attention as an alternative to achieve low power and high-bandwidth interconnects in the multi-/many-core era [2].

Nevertheless, the design of energy-efficient photonic NoCs faces many new challenges. Unlike electrical NoCs, static power dominates the overall photonic NoC power budget (e.g., 75% reported in [3]). Worse, the energy conversion efficiency of the laser sources is low (e.g., 50% reported in [4]), which further aggravates the total power loss. While the static power of photonic NoCs is fixed owing to the predetermined network design and the constant laser source injection, the network traffic manifests substantial runtime variation [5, 6]. When the traffic is below the provisioned network bandwidth, the NoCs will manifest a significant static power overhead. Moreover, due to the lack of optical logic gates and storage, existing photonic NoC routing approaches are either static or relying on additional components (such as duplicated optical networks [8] and electrical buffers [7]) to achieve adaptivity. These methodologies, however, fail to exploit existing photonic network resource effectively and increase the overall NoC latency and power due to the inclusion of auxiliary components. In summary, the emergence of photonic NoCs calls for a new set of techniques to optimize their energy efficiency. To this end, we propose ESPN, an energy-star photonic NoC architecture. Specifically, we make the following contributions: (1) We propose a dynamic photonic NoC design that allows network resources to adapt with run-time traffic characteristics. In our design, the network resources are partitioned and supplied with separate laser sources to enable dynamic network resource management strategies via traffic-aware bandwidth provisioning. (2) We propose all-optical adaptive routing to accelerate data communication. Our adaptive routing leverages low-latency optical links to establish data paths and thus avoids introducing high latency and power hungry auxiliary routing components.

The rest of this paper is organized as follows: Section 2 proposes the ESPN architecture design. Section 3 describes our experimental methodology. Section 4 describes the design of all-optical routing. Section 5 evaluates the benefits of ESPN and Section 6 concludes the paper.

2. The Proposed ESPN Architecture

2.1 An Overview of ESPN

In this section we provide an overview of the proposed ESPN architecture design. ESPN is an architecture targeting future high-throughput systems, so our exploration and evaluation targets ITRS 22nm technology [9]. ESPN consists of one multi-processor chip and two laser source chips connected by off-chip optical fibers and electrical wires on the PCB. The multi-processor chip consists of three vertically stacked dies using 3D packaging technology [10]. The processor and caches die contains processor cores, private L1/L2 caches and electrical routers. The control die, which operates as the interface between the processor and caches die and the optical die, integrates driving circuits, sense amplifiers, and control circuits for the optical components (e.g. the ON/OFF switch of turn resonators and modulators/photo-detectors). The optical die, which integrates the waveguides and ring resonators, is connected to the control die using Through-Silicon Vias (TSVs) [11]. These optical components are built using CMOS-compatible monolithic integration to reduce cost.

ESPN employs a 2D mesh dynamic optical network which consists of several sub-networks to support traffic-aware dynamic network resource allocation through real-time tuning of the laser sources; each of the sub-network is supplied with individual external laser source. The laser lights are provided by external Vertical-Cavity Surface-Emitting Laser (VCSEL) source chip [13] and coupled into the multi-processor chip via off-chip fibers. The laser lights are separately conducted to the waveguides in horizontal and vertical directions through on-chip splitters. A basic switch element of ESPN consists of an electrical router and an optical switch. The electrical router surrounds a processor core and locates on the processor and cache die; the optical switch is located on the optical die. In addition, ESPN uses all optical adaptive routing to further improve the energy efficiency of photonic NoCs.

2.2 Dynamic Resource Allocation in Photonic Network

ESPN achieves dynamic resource allocation by partitioning the interconnection network into multiple sub-networks. Each sub-network provides a fraction of the aggregated bandwidth
and is driven by separate lasers. The bit widths (i.e., wavelengths) of the data channels are divided among the sub-networks. The sub-networks can be dynamically activated/deactivated based on the run-time bandwidth estimation. To further minimize the power consumption of an inactive sub-network, the driving circuitries, and heaters (we assume each heater is dedicated to one turn resonator or shared by one modulator/photo-detector array [14]), along with the photonic components are turned off.

2.4. The Architecture Support

Both the optical switch and the network interface in conventional optical network need to be modified to support the dynamically partitioned photonic network. In ESPN, each electrical router is connected to a processor with its private L1/L2 caches, as shown in Figure 1(a). The routers and processors are located on the processor and cache die, atop the corresponding optical switches. The electrical router and the optical switch communicate with each other through TSVs and optical/electrical signal converters on the control die. Each electrical router contains four input and four output interfaces. Each output interface is shared by two output directions; while an input interface is dedicated to an input direction to support all-optical adaptive routing.

(1) Network Interface: As shown in Figure 1(b), the output interface uses a message FIFO to distribute incoming messages (from local processor) among the active sub-networks. Message transmitters are deployed to support optical adaptive routing, each of which serves two sub-networks in different output directions. A message transmitter is available when not transmitting data and its connected sub-networks are active. The sub-network selector assigns messages to the available transmitters in a round-robin fashion. When deactivating sub-network(s), ESPN employs two mechanisms to avoid destroying in-flight messages. First, the laser sources of the deactivated sub-network(s) remain on for $T_{\text{MAX}}$ cycles in order to complete the transmission of traversing messages in network, where $T_{\text{MAX}}$ is the round-trip transmission cycle between two nodes with the maximal distance (We evaluated $T_{\text{MAX}} = 3$ in 8×8 network in this study). Second, a message in message transmitter will not be eliminated immediately after being sent to the network owing to potential re-transmission. The message is destroyed only after reaching its destination. Similar to the output interface, the input interface employs a central FIFO to buffer traffic from active sub-networks.

(2) Central Controller: The state of sub-networks is determined by the network pressure, which indicates the ratio of communication demand and available bandwidth. In our design, the network pressure of each output interface is estimated as $2^{\left\lfloor \frac{n-n_0}{\text{b}_{\text{ps}}} \right\rfloor}$, where $n$ is its FIFO data count and $n_0$ is the FIFO capacity. This allows our design to better adapt with bursty and un-predictable NoC traffic such as the hotspots. The output interfaces periodically send their network pressure to the central controller. The central controller employs an electrical H-tree network composed of differential transmission lines (T-line) [21] located on the control die to collect network pressures from all routers. This H-tree network is similar to the common H-tree clock distribution network but with reverse signal flow and light load. The central controller aggregates the network pressure, and then identifies the number of active sub-networks (via the Network Status Lookup Table), expressed as
\[
\begin{cases}
m \leq p \leq p_{\text{max}} & \quad \text{if } m \leq p_{\text{max}} \\
p_{\text{max}} & \quad \text{if } m > p_{\text{max}}
\end{cases}
\]
where \( m \) is the total number of sub-networks, \( p \) is the current network pressure and \( p_{\text{max}} \) is the threshold to activate all sub-networks. The laser controller generates the laser source control information based on the required number of active sub-networks. According to our simulation results, we choose 512 for ESPN (F-1-2), ESPN (F-2-1), ESPN (P-1-2), ESPN (P-2-1), and 640 for ESPN (F-2-2) and ESPN (P-2-2) (Please see Section 4.1 for the definitions).

When the network pressure fluctuates, the subnetwork status should be adjusted correspondingly. The latency of sub-network activation/deactivation consists of the following components: delay in the H-tree network, delay in the central controller, delay from the central controller to laser, and the laser operation delay. Assuming a 3cm width waveguide, the H-tree differential T-line delay in 22nm is estimated as 8.04 ps/mm [22]. So the H-tree transmission delay is estimated to be 240 ps. The delay in the central controller is assumed to be 3 cycles (600 ps under 5GHz clock). The delay to the laser source chips is determined by the distance between the laser source chip and the processor chip and is assumed as 200 ps. The laser operation delay is 10 ps to 100 ps as mentioned in Section 2.3 and we use 50 ps in our simulations.

3. All Optical Adaptive Routing

Although our dynamic resource allocation reduces network power considerably, it could incur performance degradation due to the reduced network bandwidth. The adaptive routing achieves load balance and could compensate for the reduced network bandwidth; nevertheless existing optical-based routing algorithms [24] are mostly static due to the inherent buffer-less nature of photonic NoCs. The use of auxiliary electrical network routing [2] introduces additional hardware and performance overhead. To overcome these limitations, we propose an all-optical adaptive routing scheme by leveraging the low-latency optical network to route messages. To our knowledge, this is the first work that explores optical adaptive routing in mesh network without relying on the high-power electrical components.

![Figure 2(a): The request signal in routing examination and forwarding](image)

Our proposed all-optical adaptive routing first establishes an optical circuit switch path between the source and destination node and then transmits messages via that path. To achieve good tradeoff between routing complexity and network performance, we adopt minimal adaptive shortest-distance routing algorithm [24], which searches for the optimal routing path among all the shortest distance paths. The path establishment consists of the following scenarios: (A) The source node sends request signals along the shortest path(s) to check their availability. While proceeding, a request signal reserves photonic links along the path for the upcoming message. (B) In case the request signal encounters a blocked link and fails to reach the destination, the signal carrying the information of the blocked position is transmitted back along its reverse path and releases the reserved links. (C) If the request signal reaches its destination, the signal that indicates successful link acquisition is transmitted back to the source node. The message transmission then starts along the reserved links. (D) If all the request signals are blocked and fail to reach the destination, the source node retries the next path. Each scenario is described below in detail.

![Figure 2(b): An example of blocked link in adaptive routing](image)
3.1 Scenario 1 - The Traversal of Request Signal

The request signal travels along the request channel, which is part of the routing channel. The routing channel also contains the response channel driven by reverse laser light. The request channel consists of an even number of waveguides (two waveguides are shown in Figure 2(a)). The wavelengths of the request signal are organized as two groups: Path Hops (PHOP) and Request Hops (RHOP). The PHOP stripes across the two waveguides and is divided into several sections (PHOP1 to PHOPn), which sequentially records the routing information in n hops. Each PHOPx consists of four optical bits (wavelengths), which represent four possible turn directions. At each switch, if the downstream switch is available, the active PHOP bit drives the turn resonator of the corresponding direction to route the request signal and upcoming message.

Figure 2(a) illustrates a case in which the request signal traverses through three hops (go straight, turn left, and finally received by the local node). Each switch snoops on the PHOP1 wavelengths to detect the turn direction of request signal. PHOP1 is eliminated after the signal is routed through the current switch. So PHOP2 needs to be moved to PHOP1 to be detected by the next hop. As a result, all PHOP2 needs to be moved to PHOP1. To achieve this, we apply either physical shift or frequency translation mechanism proposed in [7] to one waveguide, which respectively moves the optical bits to the same or different wavelengths at another waveguide. For example, Figure 2(a) shows that in hop 1, PHOP1.3 are frequency translated to PHOP2.4 (different wavelengths) while PHOP2.4, 6 are physically shifted to PHOP1.3, 5 (the same wavelengths). Our proposed design implements the all-optical adaptive routing by leveraging the RHOP signals and response channel rather than static X-Y routing in [7].

3.2 Scenario 2 - Path Availability Identification

After sending the request signal, the source node needs to be notified on path availability. If the requested path is currently blocked, the source node will be notified about the block position and then plan an alternative path, which is achieved using RHOP and REPLY. The RHOP is part of the request signal and is duplicated across the request waveguides. An active RHOP1 indicates that the current distance to the destination node is i-1 hops. RHOP is decreased at each hop by frequency translating the activated bit to the next one. For example, in Figure 2(a), in each hop the RHOP1 is eliminated and RHOP2, 6 is frequency translated to RHOP1, 5 in the other waveguide. When the request signal encounters a blocked link or reaches its destination, the REPLY is modulated by physical shifting the RHOP signal from the request waveguides to the response channel and then transmitted back to the source node. The source node examines the REPLY to decide whether the path has been successfully established.

Figure 2(b) illustrates an example in the context of a 4×2 2D mesh topology. In this example, node 1 needs to send a message to node 7, while the link between nodes 3 and 7 is currently blocked. The source node simultaneously generates up to two request signals (on a minimal adaptive routing basis along both coordinates) to accelerate link establishment. In this example, node 1 generates two request signals to its South (request A) and East (request B) output ports. The request A is frequency translated from RHOP3 to RHOP2 at node 2, indicating the message proceeds. Due to the blocked link, node 3 modulates the REPLY by physical shifting RHOP to response channel and then eliminates request A. The request B successfully reaches its destination, node 7. Thus node 1 receives two REPLYs from the two response channels and identifies the REPLY from the East port as a successful link establishment. During the transmission of request signal, in case that several request signals contend for one output port, the highest priority will be given to the one that is closest to its destination by activating the corresponding RHOP signals. Such distance-class ordering mechanism ensures a deadlock-free network [24].

3.3 Scenario 3 - Data Transmission

The source node starts transmitting data along the reserved links once the REPLY indicates that a path has been established. When the data transmission completes, an optical pulse traverses back along the path to tear down the links.

3.4 Scenario 4 - Alternative Path Selection

In case that all the request signals are blocked and fail to reach the destination, the source node will plan an alternative path. The source node first examines the returned REPLY and locates the blocked position. It then retries the next path using the unblocked links plus a detour for the blocked links. This path is still one of the shortest-distance paths between the source and the destination. For example, in Figure 2(b), if both requests fail to reach the destination, a possible alternative to request A is links 1-2, 2-6 and 6-7.

3.5 The Path Request Latency

In all-optical adaptive routing the path request latency is crucial since the source node may attempt to establish a link multiple times before succeeding. This latency is characterized by the source resonator modulation latency, interim resonator drive latency, destination resonator modulation latency, and optical link latency. For each hop, the request signals fall into one of the three cases: (a) frequency translated and then forwarded to the next hop (passed), (b) transmitted back to the source node through the response channel (blocked), or (c) received by the local node (received). In case (a), the request signal needs to be received and driven to the corresponding resonator, the same as in case (b), except that the to-be-driven resonator is at the response channel rather than the request channel. In case (c), the request signal drives resonators to eliminate itself. In all cases, the latency is determined by three factors: (1) the time to receive PHOP and RHOP signals from the request channel, (2) the latency of performing a single-level CMOS logic to establish a path, and (3) the time for driving the physically shifted and frequency translated resonators. We use the latency parameters from [9, 11, 28] to estimate the request attempt round-trip delay. The optical network, including resonators and peripheral circuitry, operates at 5 GHz. Our calculation shows that the high-speed optical signal takes ~3.5 ns for the request signal to make a 14-hop (i.e. the maximal hops in a 8 × 8 mesh network) round trip.

4. Experimental Methodology

4.1 Machine Configuration and Workloads

Our evaluation is performed using a simulator developed from Simics/GEMS framework. Simics [25] is a full-system, functional simulation framework whereas GEMS [26] provides...
cycle-accurate timing for multiprocessor memory systems. We used GARNET [27], a detailed cycle-accurate on-chip network model incorporated within the GEMS framework, and extended it to support our proposed optical NoC architecture. All simulations are performed on an 8×8 mesh network as listed in Table 1. We explore different sub-network partition schemes while keeping the data channel bandwidth, i.e. the product of the number of wavelengths per waveguide and the number of bundled waveguides per data channel, constant. The baseline case, “E-deterministic”, includes an auxiliary electrical network configured as in [2]. Among all studied design alternatives, the “O-Deterministic” uses the optical channels and X-Y static routing to establish path, similar to [7]. The “O-Adaptive” adopts all-optical adaptive routing. The “ESPN (F-m-n)” and “ESPN (P-m-n)” are fully splitting and data channel splitting ESPN respectively, where each waveguide is 64/m bits wide and with 8/n waveguides. Our modeled system consists of 64 processing cores with private L1 and L2 caches fabricated using 22nm processing technology. We assume the interconnect network clock is 5GHz with a supply voltage of 0.5V. We used 128-state MMP synthetic traffic (i.e. Bit-compliment, Tornado, Bit-reverse, and Random) in our simulations. The MMP synthetic traffic generates a time-varying NoC utilization by modulating the rate of a Bernoulli injection process on the states of a Markov chain [24]. The injected messages in MMP synthetic traffic are 64 bytes (the cache line size) and 8 bytes (the invalidation message size) respectively.

Table 1. The evaluated NoC design

<table>
<thead>
<tr>
<th>NoC Design</th>
<th>Link Establishment</th>
<th>Routing Scheme</th>
<th>Network Division</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-Deterministic</td>
<td>Electrical</td>
<td>X-Y Static</td>
<td>-</td>
</tr>
<tr>
<td>O-Deterministic</td>
<td>Optical</td>
<td>X-Y Static</td>
<td>-</td>
</tr>
<tr>
<td>O-Adaptive</td>
<td>Optical</td>
<td>Adaptive</td>
<td>-</td>
</tr>
<tr>
<td>ESPN (F)</td>
<td>Optical</td>
<td>Adaptive</td>
<td>Fully Splitting</td>
</tr>
<tr>
<td>ESPN (P)</td>
<td>Optical</td>
<td>Adaptive</td>
<td>Data Splitting</td>
</tr>
</tbody>
</table>

4.2. Power Estimation Methodology  

Table 2. Optical loss in various components

<table>
<thead>
<tr>
<th>Component</th>
<th>Optical Loss (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical coupler</td>
<td>1 dB</td>
</tr>
<tr>
<td>Interlayer coupling loss</td>
<td>1 dB</td>
</tr>
<tr>
<td>Filter drop</td>
<td>1 dB</td>
</tr>
<tr>
<td>Waveguide loss</td>
<td>1.3 dB/cm</td>
</tr>
<tr>
<td>Bending loss</td>
<td>0.5 dB</td>
</tr>
<tr>
<td>Non-linear loss</td>
<td>1 dB</td>
</tr>
</tbody>
</table>

We used the statistics reported in [30, 31] for the optical network power estimation, as summarized in Table 2. The electrical energy coupling efficiency of the laser source ranges from 30% [31] to 50% [30, 32]. We used the median value of 40% in this study. Another important factor of the total power consumption is the required optical detection power for photodetectors, which is related to the expected Bit Error Rate (BER). We adopted BER of $10^{-15}$ in our study and [12] shows that each photo-detector requires at least 5 μW power under 5Gb/s modulation rate. By default, all turn resonators are set to OFF state and tuning energy is required when switching to ON state [29]. This energy is assumed to be 100 fJ/bit [18]. Besides, the power consumed by each modulator is approximately 200 fJ/bit using advanced driver circuits with poly-Si carrier lifetimes of 0.1-1 ns and modulation speed of 5 Gb/s [18].

In a typical photonic NoC with auxiliary electrical network [2], the total power consumption is the sum of power dissipated by both optical and electrical networks. The power consumption of the electrical network is modeled based on [17, 23], which assumes the energy required to transmit one bit under 22nm technology is 0.83 pJ plus 0.34 pJ/mm link power.

5. Evaluation

In the evaluation section, we explore the design spaces of ESPN and evaluate the performance and power benefits of the proposed techniques. Since the main purpose of this study is to optimize the existing optical network, in this section we mainly compare ESPN design with conventional mesh-style optical network [2, 7].

5.1 Network Latency

Figure 3 shows the network latency under 128-state MMP synthetic traffic. E-deterministic and O-deterministic exhibit the worst performance owing to the deterministic routing. O-adaptive and ESPN (P) benefit from adaptive routing and thus improve performance by 20%-25%. However, due to the reduced bandwidth and subnetwork switch delay, ESPN (P) incurs 1%-3% performance degradation compared to O-adaptive. ESPN (F) gains a 5%-10% performance improvement over ESPN (P) by deploying dedicated routing channels in each sub-network.

5.2 Power and Energy Efficiency

Figure 4 shows power breakdown of the investigated NoC design. We observe that among all sub-network partitions, ESPN (P-2-2) shows the best power efficiency, which yields 50% savings on average compared to E-deterministic. When the number of wavelengths within each waveguide drops from 64 to 32, an additional 10% power saving is observed due to the alleviated optical coupling loss in modulators and photodetectors. On ESPN (F), the power for the routing channels increases due to the deployment of dedicated resources in each sub-network. ESPN (F-2-2) reduces power by 46% compared to that of E-deterministic. The power savings vary since the average number of hops that a message traverses in the 8×8 network varies with traffic patterns. Figure 4 also shows the energy consumed per message. As can be seen, ESPN (P-2-2) and ESPN (F-2-2) save 57% and 58% energy respectively compared to the
baseline case. The above two network configurations have different performance and power characteristics but exhibit similar energy per message profile.

6. Conclusions
Photonic NoCs are proposed to eliminate the latency and bandwidth bottlenecks in conventional electrical networks. However, substantial static power consumption limits the scalability of photonic NoCs. In this paper, we presented ESPN, an energy-efficient photonic NoC architecture that enables traffic-aware dynamic allocation of network resources. We showed that dynamical coupling of communication resources and network traffic could save up to 50% of the total optical NoC power. ESPN employs an optical-based adaptive routing circuit switch, which reduces the application execution time by 22%. By integrating these technologies, ESPN reduces power and energy consumption by 50% and 58% respectively on MMP synthetic traffic.

7. Acknowledgement
This work is supported in part by NSF grants 1117261, 0937869, 0916384, 0845721(CAREER), 0834288, 0811611, 0720476, by SRC grants 2008-HJ-1798, 2007-RJ-1651G, by Microsoft Research Trustworthy Computing, Safe and Scalable Multi-core Computing Awards, by NASA/Florida Space Grant Consortium FSREGP Award 16296041-Y, and by three IBM Faculty Awards. Zhongqi Li is also supported by a University of Florida Graduate Fellowship.

8. References