DCOF - An Arbitration Free Directly Connected Optical Fabric

Christopher Nitta, *Member, IEEE*, Matthew Farrens, *Member, IEEE*, and Venkatesh Akella, *Member, IEEE*

***Abstract*— In this paper we investigate the unique potential of optics to provide a family of arbitration free topologies that are not realizable using conventional electronics. This is accomplished by creating a directly connected *fabric* of waveguides that can be configured to support everything from a crossbar to fully connected topologies. The large number of waveguides required to create a Directly Connected Optical Fabric (DCOF) can be built by taking advantage of multiple photonic layers connected with photonic vias, allowing the architect to choose the degree of simultaneous communication (a parameter called *k*) necessary to meet the performance requirements and available power budget.**

In order to evaluate DCOF we developed a detailed implementation model for three different network instantiations - a crossbar similar to Corona, DCOF configured as a crossbar, and DCOF configured as a fully connected network. We analyzed the power consumption and performance of these topologies on a variety of benchmarks, including SPLASH-2 and synthetic traces. Our results demonstrate that the overhead required by arbitration is non-trivial, especially at high loads. Eliminating the need for arbitration, sizing the buffers carefully and retransmitting lost packets when there is contention results in a significant reduction in average packet latency without additional power overhead. We also show that when configured as a crossbar DCOF is the most energy efficient while maintaining excellent performance, and when configured as a fully connected network provides the best performance, but at a potentially prohibitive photonic power cost.

***Index Terms*— Network-on-chip, nanophotonic, arbitration-free, interconnect architecture**

I. Introduction

Future processors will have multiple cores on each die [1], and these multicore processors will require high bandwidth, reliable communication networks. Electrical networks are not likely to scale to large numbers of processors well (primarily for latency and power consumption reasons) - however, optical networks [2] feature less signal crosstalk, lower power loss, and higher switching speeds than electrical networks, making them ideal candidates for use in future large scale chip-level multiprocessors. For example, the authors in [3] estimate energy efficiencies in the tens of fJ/bit as being possible for nanophotonic interconnects, while Miller in [4] projects that global electrical interconnects will require at best two orders of magnitude more energy per bit.

This has led computer architects to begin to look in earnest at nanophotonics, with much of the research focus thus far on understanding and fabricating the basic building blocks, such as resonators and waveguides [3], [5]–[10]. As our understanding has grown, architects have begun to study how best to use optics in computing devices. Researchers from HP [11], Cornell [12], [13], Northwestern [14], Columbia [15]–[17] and MIT [18], [19] have proposed different topologies and arbitration/flow control schemes for such networks. The details of each design vary, but what they all have in common is that photons are generated by an off-chip laser and routed around the chip through silicon waveguides using anywhere from a few hundred to hundreds of thousands of photonic resonators. However, there has been little or no attempt to exploit the unique properties of optics - the networks proposed have simply been conventional/practical topologies (crossbars, WDM-based buses, etc.) implemented using photonics instead of electronics.

This is unfortunate, because optical networks have singular capabilities that designers can exploit to provide network configurations unrealizable using conventional techniques. For example, flat networks and directly connected fabrics are extremely attractive because they make programming tasks much easier, particularly when there are a large number of processors involved - the programmer does not have to worry about data location, which often depends on the (frequently dynamic) communication patterns in the underlying algorithm/application [20], [21]. And anything that can be done to make parallel programming simpler is of great value to the entire community.

This was a point of emphasis at the ”Future of Computing Performance” [1] symposium. Writing parallel programs is hard, and since all future processors are going to be parallel processors, it is crucial that researchers explore techniques that ease the burden of parallel program creation. The scientific algorithms community has observed that the diversity in communication cost (due to the topology and type of interconnect used) is something that makes the programmer’s task particularly difficult, and there is a push towards developing ”communication avoiding algorithms” or ”reduced communication algorithms” (Demmel et. al UC Berkeley [22]).

However, one does not have to avoid communication if the cost is low enough, and the the benefits of fully connected topologies are well known. They offer the highest bisection bandwidth, potentially the lowest communication cost, and are more resilient to failures on links, as packets can be routed through unaffected nodes. In addition, networks that have a dedicated link between the sender and the receiver do not require arbitration, only flow control – this is a significant advantage, because arbitration is an overhead that must be

C. Nitta, M. Farrens, and V. Akella are with the University of California at Davis.

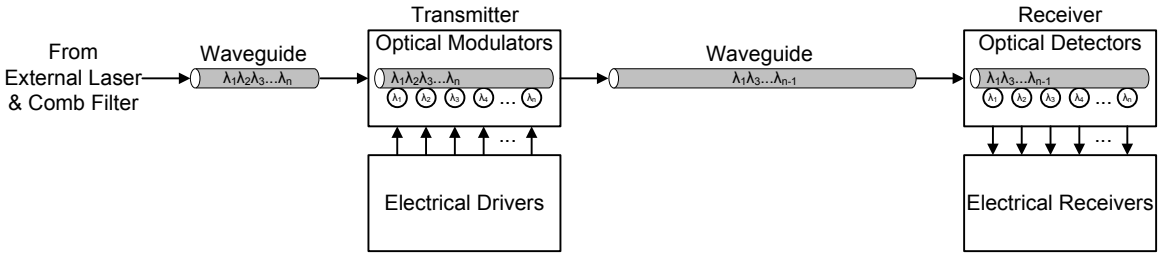


Fig. 1: Example of an Optical Link. The wavelengths are generated by an external laser and a comb filter. Upon entering enter the chip, the wavelengths enter the transmitter section, which is responsible for blocking/passing the wavelengths necessary to send the appropriate pattern down the waveguide to the receiver.

paid for all communication, while flow control only occurs when the network is becoming overwhelmed.

Electrical implementations of fully connected topologies for large numbers of processors are impractical, because of the wiring complexity and power consumption required. However, on-chip photonics can be used to create these kinds of networks, for a variety of reasons - long distance communication is possible without the need for area/power hungry repeaters, the use of dense wavelength division multiplexing (WDM) allows multiple bits of information to be transmitted simultaneously over the same waveguide (the optical equivalent of a single wire), waveguides can intersect one another without the signal being destroyed, etc. Thus, the unique properties of on-chip photonics enable the creation of networks that are highly desirable (because of the potential to ease the burden of writing parallel programs, as well as the potential performance and resilience impacts), but impossible/impractical to build using only electronics.

In this paper we introduce a family of networks that are based on a directly connected *fabric* of waveguides that can be configured to support everything from a crossbar to fully connected topologies. We refer to this fabric as DCOF (Directly Connected Optical Fabric), and it has at its core a direct optical link between every pair of nodes. The large amount of laser power needed to support simultaneous communication over all the links may be prohibitively high if the number of nodes is large, in which case it may be preferable to configure the fabric as a crossbar, or perhaps as some other topology that allows simultaneous communication between a subset of the nodes. In this paper we present an overview of how photonics work, a description of the tools we have developed, a detailed implementation model for the fabric and an analysis of the power consumption and performance when configured as the two endpoints (crossbar and fully connected). The analysis is done using simulation on a variety of synthetic and SPLASH-2 benchmarks. We also discuss different topologies that can be mapped onto a directly connected fabric and how these topologies will affect performance, power consumption and programmability.

II. Background

Figure 1 presents a typical on-chip optical link that uses an external laser as a light source. The external laser passes through a comb filter [23], which creates the necessary set of wavelengths used for communication, and then enters the chip. These wavelengths are delivered to the transmitter section of the source node via an optical waveguide.

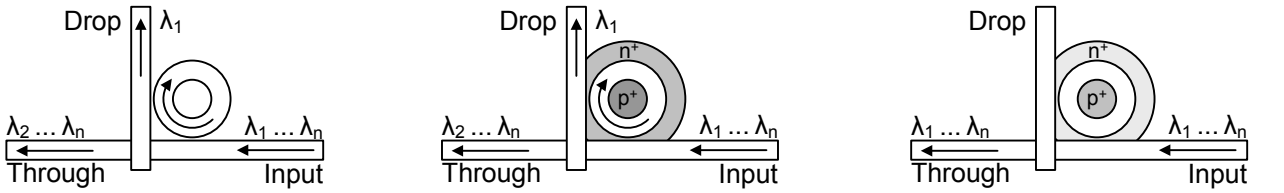
The transmitter, consisting of electrical drivers and optical modulators, uses the modulators to remove certain wavelengths (in this case λ_2 and λ_n), creating the desired pattern. This pattern then travels down the waveguide from the source to the destination node. When the transmitted value arrives at the destination, the optical detectors convert the photonic power back to an electrical signal and the transmission is complete. (One of the wavelengths serves as a clock signal, so the destination can distinguish between all zeros and no communication.) The rest of this section provides more details of how the individual components of the optical link function.

A. Basic Elements of Photonic Design

a) Resonators: Microring resonators are designed to resonate when presented with specific individual wavelengths and remain quiescent at all other times. The ability to respond to specific wavelengths enables the removal (filtering) of specific wavelengths from a waveguide, and these resonators are the primary technology used to bundle the high quantity of wavelengths per waveguide needed for Dense Wavelength Division Multiplexing (DWDM). This filtering can be achieved using either passive or active microrings. Figure 2(a) shows a high-level view of a passive microring that is biased during fabrication to extract only λ_1 from the incoming waveguide and steer it down a perpendicular waveguide.

Since the passive microrings are biased during fabrication to always respond to a single wavelength, they cannot be used for modulation. Modulating requires an active microring resonator, which is designed to change its resonance frequency based on the amount of current present in the n^+ base. Figures 2(b) and 2(c) illustrate an active microring resonator in the "On" and "Off" states, respectively. If the electrical current is present ("On" state), λ_1 is extracted from the *input/through* waveguide and sent down the *drop* waveguide – if there is no current applied ("Off" state), λ_1 will continue down the *input/through* waveguide unaffected.

Generally, it is assumed that the presence of a wavelength represents a logic 1 and the absence represents a logic 0, and the method by which an active microring modulates depends upon the configuration of the incoming and outgoing waveguides. For example, if the incoming waveguide is also the outgoing waveguide, then a zero can be created by using



(a) Passive Microring Resonator, tuned to resonate to λ_1

(b) Active Microring Modulator, which is "On" for λ_1

(c) Modulator Active Microring Modulator, turned "Off" for λ_1

Fig. 2: Microring Resonators. (a) shows a passive microring, which at fabrication time was set to resonate only to λ_1. (b) and (c) show active microrings, which use the presence or absence of charge in the n^+ base to change the wavelength they will resonate to (λ_1, here.)

the microring to remove the wavelength by bending it onto a dead end drop waveguide, and a one is created by allowing the wavelength to pass unaffected (shown in Figure 2(b)). If the incoming and outgoing waveguides are not the same, then ones are created by bending the wavelength onto the outgoing waveguide, and zeros by allowing the wavelength to continue unperturbed along the incoming waveguide. (This is shown in Figure 2(b) if the *drop* waveguide is the outgoing waveguide, and not a dead-end drop.)

b) Photonic Vias: Waveguides carrying different signals can intersect on the same layer without complete signal interference, unlike wires carrying electronic signals. Intersections of waveguides at 90 degrees allow for signals traveling down each waveguide to continue on intact, although each signal will suffer a small attenuation (often modeled as ∼0.1dB). This characteristic of photonics allows on-chip optical networks to be laid out on a single layer without the need to transition to waveguides on other layers. However, the cumulative effect of a large number of intersections may make a single layer waveguide layout infeasible – if this is the case, waveguides will need to be routed on different layers to avoid excessive intersections. In [24] we show how waveguides can be fabricated on different layers.

In the electronic domain signals can easily move from layer to layer using vias - transitioning photonic signals to different layers is done in a similar manner. Grating couplers are used to couple optical fibers and waveguides [25], [26], and it is possible to use a vertical grating coupler to connect waveguides on different layers. In our work we assume that the signal attenuation of such a coupling is 1dB, a conservative estimate considering optical fiber and waveguide couplings of less than 1dB loss have already been demonstrated.

Grating couplers are not the only possible structure for use as a photonic via. Plasmonics have the capability to drastically change the direction of light, which could be useful when changing layers; however, plasmonics suffer from high path attenuation (typically ∼0.2dB/μm [27]). Over the relatively short distances required for an inter-layer via (assumed less than 10μm), the loss experienced by a plasmonic based photonic via may be acceptable; in this work we did not use plasmonics as a photonic via, but it is an example of one possible alternative to the use of grating couplers.

B. Challenges in Photonics

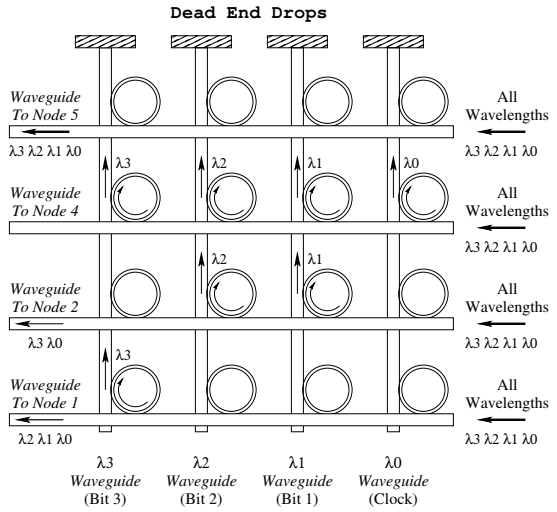
There are a number of challenges to creating large functional optical networks. For example, the wavelengths that individual microrings respond to are set during fabrication - however, variations in fabrication tolerances may require that certain microrings have their resonance frequency moved up or down slightly. Furthermore, microring resonators are very sensitive to temperature and drift spectrally approximately 0.09nm/°C. Our work thus far has shown that even though photons are being pumped into the chip by the laser, the network *itself* is thermally stable [28]. We determined this via the extensive use of our advanced simulation tool known as Mintaka [24], which calculates everything from the thermal behavior of a given photonic network to the amount of photonic and electrical energy consumed, based on the amount of network traffic that is occurring.

The resonance frequency of a ring can also be "trimmed" to account for both fabrication imperfections and thermal drift - this can be accomplished dynamically by electrically injecting current (to shift the resonance towards the blue) or by heating the ring (to shift towards the red) [3]. However, these active trimming techniques can result in a dramatic increase in the overall power requirements and even thermal runaway, as we showed in [28]. Fortunately, our simulations indicate that rings thermally drift as a group, so we have developed a technique known as a Sliding Ring Window which (when used in conjunction with less thermally sensitive PMMA clad rings [29]) can allow the designer to keep the rings within a defined Thermal Control Window [28]. In the work presented here we are assuming current injection-based active trimming of microrings with a thermal sensitivity of 1pm/°C and a Temperature Control Window of 20°C.

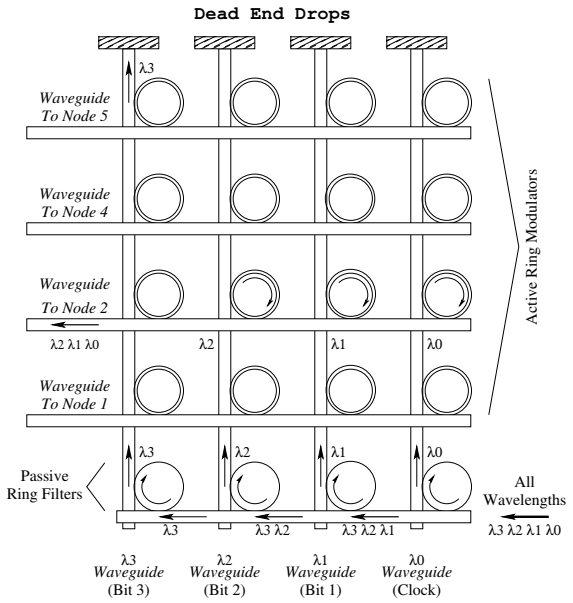
We also have examined reliability in optical links [30], and the sum total of our investigations have convinced us that the issues of fabrication tolerances and thermal drift can be overcome, and that large scale microring based on-chip networks are feasible. This led us to examine the potential benefits and costs of DCOF, which will be presented in the rest of this paper.

III. Directly Connected Optical Fabric (DCOF)

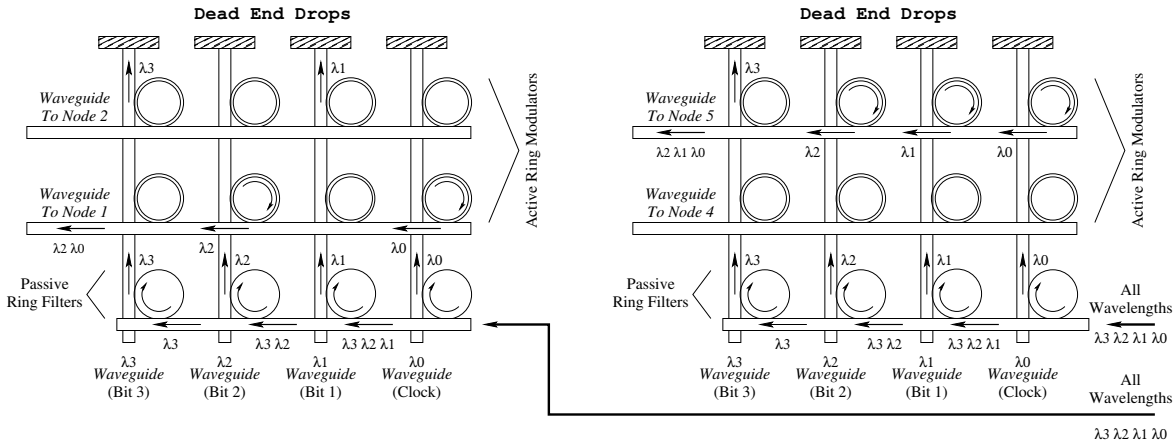
As mentioned in the introduction, one of the unique capabilities of optical networks is the ability to provide a dedicated optical link between each compute node, creating



(a) DCOF Transmit Section of node 3 in a 5 node fully connected ($k = N$-1) configuration, communicating using 3 wavelengths. All rings are actively modulated. In this figure a binary 011 is being sent to node 1, a 100 to node 2, nothing at all to node 4, and a 111 to node 5.



(b) DCOF Transmit section of node 3 in a 5 node crossbar ($k = 1$) configuration, communicating using 3 wavelengths. In this figure a binary 011 is being sent to node 2.



(c) DCOF Transmit Section of node 3 in a 5 node configuration where k=2. Note the similarity between this figure and figure 3(b). In this figure a 010 is being sent to node 1, and a 011 is being sent to node 5.

Fig. 3: DCOF Transmit Section for three different values of k: $k = N$-1 (a), $k = 1$ (b), and $k = 2$ (c).

a directly connected optical fabric - something that is not feasible in the electrical realm since the wiring area would require a prohibitive amount of both space and energy. This optical fabric requires N^2 waveguides, which is relatively straightforward to provide for modest values of N (say up to $N = 16$.) A flat network of up to 64 nodes is possible, and we show in [24] that it is possible to create a hierarchy of optical networks in order to scale the number of processors even further.

In DCOF the transmitter section of each node is modified in order to limit the number of destination nodes (denoted by k) that can simultaneously have information sent to them. k can vary between 1 and N-1, where N is the total number of compute nodes. If $k = N$-1, DCOF is operating as a fully connected network (shown in Figure 3(a)), while if $k = 1$ a crossbar topology has been instantiated upon it (shown in Figure 3(b)). The transmit section of a DCOF node with $k = 2$ is shown in Figure 3(c).[1]

The transmit section modifications consist of changing the way the photons are steered from the incoming laser to the target node. In the $k = N$-1 end case, all wavelengths of light are provided to each transmission waveguide, and a zero is created by removing a specific wavelength. In contrast, when $k = 1$, the transmitter design features a single power waveguide with a set of passive microring resonator filters, which split off each wavelength onto a dedicated waveguide. Active rings are then used to bend light from the dedicated waveguides onto

[1]Note that setting $k = 2$ does not mean a given node can communicate with any other arbitrary node - as can be seen in Figure 3(c), the receiving nodes must be put into groups so that the clock signal can be distributed properly.

the transmission waveguide in order to transmit a one. The two different cases are shown in Figure 3(a) and Figure 3(b).

The addition of the filtering microrings in order to vary k increases the link losses, due to the increase in the number of on and off-resonance rings through which the light must travel. Our analysis shows that for a 64 node DCOF with 64-bit data path, the link losses increase from 9dB to 9.7dB when k changes from 63 to 1. However, even though the losses have gone up, the photonic power required has fallen by a factor of approximately 53 (102 watts versus 1.9 watts assuming 10GHz operation) since the laser only has to be powerful enough to transmit to a single destination, not all 63.

One of the key advantages to DCOF is that since the fabric provides direct links between nodes, no arbitration will be required for any network topology that is mapped onto it (for any value of k, in other words). Not requiring arbitration is a significant advantage enjoyed by DCOF, because arbitration is a constant *pre-paid* cost (in terms of both power and performance) that is incurred every cycle, whether or not a communication takes place. Given the finite buffering that exists at each node link level flow control will be required, but flow-control only kicks in when communication is happening and the receive buffers are full, which is obviously a much less frequent event.[2] Arbitration is also a potential point of failure, since if any part of the arbitration network fails the entire system is rendered useless, making the network less resilient. And resilience is important to keep in mind when designing with new technologies such as on-chip photonic devices, whose fabrication process is not very mature.

Perhaps most significantly, arbitration requires some level of "global knowledge" of the whole system to decide who gets to communicate (i.e. share the receiver), and global structures limit scalability. By eliminating arbitration and replacing it with purely local information about the flow status of a link, DCOF allows the realization of more scalable on-chip networks, which are more resilient, have better performance and consume less power.

Flow control is accomplished in DCOF using an Automatic Repeat-reQuest (ARQ) scheme. If a flit arrives at a reception node and there is no available space in the buffer, the flit is dropped and the ACK is not sent back. A Go-Back-N ARQ scheme was chosen over a conventional credit based flow control approach since multiple flits can be in flight simultaneously on a single waveguide - or, to put it another way, the round trip of a single link can be much greater than 2 cycles. The ARQ scheme allows for efficient flow control without the need for excessive buffering. Reliable communication is another benefit of using an ARQ scheme for flow control, since lost flits or potentially corrupted flits can be retransmitted. The size of the ARQ ACK token is chosen to be 5 bits, which allows for worst case round trip propagation delay and therefore supports uninterrupted flow.

Considering the number of node connections (and hence the number of required waveguide crossings) and an assumed 0.1dB loss per intersection, a single layer implementation of DCOF will not be practical. The use of photonic vias and multiple photonic layers, though, do enable the creation of directly connected fabrics like DCOF. Since the number of waveguides needed in DCOF grows quadratically with node count, simply estimating the area necessary may be misleading - therefore, we present an entire layout for a 16 node DCOF using a 16-bit bus in [24], [31]. Assuming an 8μm ring pitch (3μm ring and 5μm ring spacing), and a 1.5μm waveguide pitch (0.5μm waveguide and 1μm waveguide spacing), the network as illustrated occupies an area of $\sim$1.15mm^2. A 64 node DCOF could be constructed by clustering four groups of 16 nodes and interconnecting them in the same way 4 node clusters are interconnected in the 16 node case. Laying out a DCOF fabric in this fashion requires that the number of layers grow as $log_2(N)$, though fewer layers could be used at a cost of more complicated waveguide routing. Given our assumed layout technique (which routes waveguides around the microring area) a 64 node DCOF with k = 1 will require $\sim$58.1mm^2. This is large, but not unreasonably so.

DCOF gives the computer architect a fabric upon which specific photonic network topologies can be mapped, and these topologies can be chosen based on the expected workload (supercomputing versus streaming, for example) and/or the specific end-market (such as consumer, server, or a blade in high-performance computing.) In each case the physical layout of DCOF remains the same, and the customization is done in a modular fashion by changing the number of filters needed to achieve the desired k (shown in Figure 3).[3] In the next section we will present an analysis of the performance and power consumption of DCOF with two different topologies instantiated upon it: when it is used as a crossbar, which we will call DCOFk1 (since k=1), and when operating as a fully connected network (which we will call DCOFk63).

IV. Description of Experimental Topologies

In order to analyze and evaluate the arbitration-free DCOFkX topologies, we needed a representative network to compare it to. We wanted to compare DCOF to a flat topology which had identical total, bi-sectional, and link bandwidth, so we created the Crossbar Optical Network (CrON). CrON is modeled closely after the Corona design, primarily because Corona has been very carefully scrutinized over the years and there are enough details publicly available to allow it to be modeled relatively accurately.

A. Crossbar Optical Network (CrON)

The Corona design is a 64 x 64 256 bit crossbar operating at 10GHz (double clocked 5GHz). Therefore, CrON also assumes 64 nodes and a similar serpentine layout to bring the waveguides to each crossbar node, although CrON uses a bus width of 64 bits instead of 256. The decision to model a 64 instead of 256 bit data path was driven by the fact that we were modeling a 64 "core" instead of a 256 "core" system. Table I highlights the structural differences between Corona and CrON.

[2]Some networks require both arbitration and flow control because of the finite buffer constraint. In some networks (such as Corona [11]) it is handled by a single mechanism.

[3]Currently k must be chosen at fabrication time, although it may be possible in the future to change the topology dynamically.

TABLE I: Corona/CrON Network Parameters

Network	Tech	WGs	Microrings		Bandwidth		
			Active	Passive	Total	Bisection	Link
Corona	17nm	257	∼1M	∼16K	20TB/s	20TB/s	320GB/s
CrON	16nm	75	∼292K	∼4K	5TB/s	5TB/s	80GB/s

Arbitration in CrON is handled in a manner similar to the Token Channel with Fast Forward described in [32]. Due to the nature of the protocol, a processor can wait up to 8 clock cycles (at 5GHz) to receive an uncontested token. Increases in die area and node count will increase the serpentine waveguide length and therefore increase propagation delay, meaning that the delay for uncontested tokens will grow with increased clocking speeds, die area, and node count. (The CrON design, however, does have the capability of a simultaneous one-to-many transmission if a single node were by chance to acquire arbitration tokens for multiple receivers.) The Token Channel with Fast Forward protocol was chosen over the Token Slot [32] since Token Slot can lead to node starvation. Token Channel with Fast Forward was also chosen over Fair Slot protocol since a broadcast waveguide is required in order to support Fair Slot [32], which our detailed simulations show increases the photonic power required for arbitration by a factor of 6.2.

B. DCOF With k=1 (DCOFk1)

As stated previously, DCOFk1 incorporates additional microring resonators in the transmitter section of each node, which are used to limit the number of destination nodes that can simultaneously have information sent to them. DCOFk1 in essence has a locally controlled demultiplexer in its transmit section, making it a many-to-one crossbar - a single node can simultaneously receive from multiple sources, but can send to only one. CrON, on the other hand, has the equivalent of a receive multiplexer which must be globally arbitrated.

Figure 3(b) shows that the DCOFk1 design is not technically limited to transmitting to a single receiver at a time; the actual limitation is that each individual wavelength can only go to one receiver. The wavelength not being sent to node 2, λ_3, could be sent to node 1, 4 or 5, for example, but not to all of them. This limit of 1 wavelength per receiver effectively prevents bus-width sized messages from being transmitted to multiple receivers, because one of the wavelengths must serve as the clock wavelength and it can only go to a single receiver at a time. It would be possible to add a clock wavelength for every grouping (e.g. 8 bits) in order to send smaller transmissions to multiple destinations, but still allow for full sized transmissions to a single destination – if a clock wavelength is used for each grouping of data bits, then the limit of simultaneous transmissions is bound by the granularity of the grouping. When k is greater than 1 there must be multiple sets of power waveguides and filtering rings, and each node in such a design would be limited to a single transmission per receiver group, but would be capable of k simultaneous transmissions. This can be seen in Figure 3(c).

Table II illustrates the structural differences between CrON and DCOFk1. Note that the number of waveguides in CrON is somewhat misleading - if one considers a single loop around the chip as just one waveguide, then the number is 75; however, if one considers each segment between nodes to be a separate waveguide then there are actually ∼4.6K, which is more than is used by DCOFk1. DCOFk1 also requires ∼88% more microrings than CrON, although there are in fact fewer *active* (power-consuming) microrings required in DCOFk1 than in CrON. As stated earlier, the total, bi-sectional, and link bandwidth of the two networks are identical.

C. DCOF With k=63 (DCOFk63)

As described in Section III, a 64 node DCOFk63 is a fully connected network. As Figure 3(a) illustrates, since each source destination has dedicated photonic power, it is possible to use fewer wavelengths for flow control. Instead of using the 5-bit ACK based ARQ flow control described previously, we can dedicate one single wavelength to function as a single flow control bit. This flow bit is responsible for conveying whether flow is enabled or disabled, and decreases the necessary wavelength count from $D+6$ (D data bits, 1 clock, and 5 ACK) to $D+2$ (D data bits, 1 clock, and 1 flow). Disabling flow is accomplished by turning off the flow bit to a given node.

There are a variety of ways the re-enabling of flow can be accomplished: all flows can be simultaneous re-enabled, flows can be re-enabled in a round robin fashion, or based on proximity, or most recently received, etc. In the experiments presented in Section VI a round robin re-enabling scheme was employed because it is fair, and will not result in large bursts of traffic from all sources during the re-enabling stage.

V. Mintaka

Performing a thorough analysis of DCOF requires a detailed simulation infrastructure for photonic networks, which we developed and call Mintaka. A detailed description of Mintaka is in [24] - we will only present a basic overview here. A list of the optical simulation constants used in Mintaka are provided in Table III; where noted the parameters are taken from the Corona published works, while all other parameters are extrapolations (within theoretical limits) from experimental prototypes fabricated at UC Davis.

The photonic power estimates in Mintaka are derived using a link loss approach, and power levels for each possible path through a link are maintained (all photonic energy is tracked inside Mintaka). Mintaka is also capable of performing a thorough thermal analysis, which is essential to understanding the true power consumption in on-chip optical networks, since

TABLE II: CrON/DCOFk1/DCOFk63 Network Parameters

Network	Tech	WGs	Microrings		Bandwidth		
			Active	Passive	Total	Bisection	Link
CrON	16nm	75	∼292K	∼4K	5TB/s	5TB/s	80GB/s
DCOFk1	16nm	∼4K	∼276K	∼280K	5TB/s	5TB/s	80GB/s
DCOFk63	16nm	∼4K	∼260K	∼260K	315TB/s	160TB/s	80GB/s

TABLE III: Simulation Optical Parameters

Description	Value	Description	Value
Waveguide		**Microring**	
Width	0.3μm	Diameter	3μm
Spacing	1μm	Spacing	5μm
Minimum Bend Radius	1.5μm	Resistance	10Ω
Attenuation*	0.3dB/cm	Capacitance	10fF
Intersection Attenuation	0.1dB	Quiescent Current*	10μA
Grating Attenuation	1dB	On Resonance Attenuation	0.5dB
Bend Attenuation	2.25e-4dB	Off Resonance Attenuation*	1.5e-3dB
Photodetector			
Width	3μm	Attenuation*	3dB
Height	0.3μm	Capacitance*	10fF

* The numbers are taken from the Corona published works [3].

items such as microring "trimming" power and buffer leakage are functions of temperature.

Many of the electrical components used in Mintaka were constructed in a manner similar to those used in ORION 1.0 [33], although electrical technology data such as transistor capacitances were taken from CACTI 6.5 [34] (for technology parameters from 90nm to 32nm), and MASTAR following ITRS 2009 [35] (for below 32nm). Unlike ORION and CACTI, Mintaka does not size transistors by scaling based on the technology point, but sizes transistors based on the required switching period and load to be driven (a safety factor is also included). Transistor folding is also accounted for once the overall transistor width is determined from the load and switching period. The wire technology data is based on [36], and the methodology used in [37] is used in Mintaka to determine the correct wire sizing (local, semi-global, or global) given the desired bandwidth and wire length.

Great care was taken in developing actual floor-plan layouts. The floor-plan for each network is integrated into the electrical/optical power/sizing calculations, since both the optical and electrical power requirements depend upon the distance which the signals must travel, and the minimum size of some sub-components is dependent upon the power requirements.

A closed loop solver was used to determine the steady state power required for each network. Since all power consumed in the network was maintained for each floor-plan component, this power and floor-plan could be input into Hot-Spot 5.0 [38]. The Hot-Spot steady state solver was used to determine the updated temperatures for the floor-plan components, which then was input back into Mintaka. Static leakage power and trimming power are functions of temperature; therefore, an updated power consumption was calculated and input back into Hot-Spot. The iterative process continued until Mintaka/Hot-Spot converged on a steady state solution.

Mintaka was validated by comparing the optical and electrical components separately. The optical validation was done by comparing its link loss calculations to those published for Corona [3], when the same input parameters were used. The electrical components were compared against intermediate values generated inside CACTI and ORION when using the same technology data. The differences observed in the electrical components were due to the transistor sizing done by Mintaka vs. the static scaling from 0.8μm assumed in ORION and CACTI.

According to Mintaka, the worst case path attenuation for CrON is 17.3dB, which is higher than the 13dB calculated for Corona in [3]. There are two reasons for the difference: first, the Token Channel arbitration used in CrON requires more off resonance rings in the worst case path than are needed using the arbitration assumed in Corona [3], and second, the CrON worst case path requires additional waveguide path length and bends to allow the token power feed to flow in the same direction as the arbitration tokens. It was discovered using our simulations (and with the help of several of the Corona authors [39]) that if power flows counter to that of the tokens, a gap in photonic power can occur when a token needs to be injected. This discovery in no way diminishes or negates the previous findings regarding photonic tokens, but does change the structures that must be assumed for token injection.

VI. Experimental Setup

In order to evaluate the performance of the different topologies we created a trace-driven network performance simulator to determine the latency, average and maximum queue depths,

average and peak bandwidth, and total execution time. In [40] we showed that not including packet dependencies can yield misleading performance results, so we took the dependency tracking simulator used in [40] and added the CrON and DCOF networks to it in order to more accurately ascertain network performance. The base architecture we modeled was a 64 node network with a 64-bit data path between nodes, built using 16nm technology. The "cores" were assumed to operate at 5GHz and be capable of generating and consuming one 128-bit flit per cycle. The on-chip network occupies an entire level of a 3D stacked processor design, with an area of 484mm^2.

The "traces" (more correctly Packet Dependency Graphs, or PDGs) used in the performance simulations were a combination of synthetic traffic patterns and SPLASH-2 benchmarks. The synthetic traffic patterns chosen were *uniform random*, *negative exponential distribution* (NED) [41], *hotspot*, and *tornado*. All synthetic traces were run with a standard range of offered load (no dependencies) in order to determine maximum network throughput and average packet/flit latency. The SPLASH-2 benchmark PDGs used were a 16 million point FFT, Water SP, LU, Radix, and Raytrace. The PDGs were obtained from multiple 64 node full system simulations using the GEMS framework that includes the Garnet network simulator; packet dependencies were then inferred using the algorithm outlined in [40].

A. Buffering Analysis

The amount and configuration of network buffering is an important factor in analyzing the performance and power consumption of on-chip networks. The amount of transmit and receive buffering (in the form of FIFOs) at a given node alone is insufficient to determine the power/performance of the network, however - for example, one cannot assume shared buffering for all transmitters at a node in CrON, since multiple flits can be simultaneously transmitted. For the buffers to be shared, one must also include an electrical crossbar to connect the buffers to the transmitters. The same is true on the receive side in DCOFk1 - sharing the receive buffer requires a crossbar to connect the receivers to the shared buffer. These local crossbars require N-1 input and output ports, and the power consumed by these crossbars reduces the power advantages of using photonics.

Fortunately, it is possible for DCOFk1 to have a smaller local crossbar, with N-1 input ports and less than N-1 output ports; this would allow the same number of flits as output ports to be simultaneously transferred from the private buffers to a shared buffer. The same is not true of the transmit side of CrON, though, since flits must be sent sequentially once arbitration has been obtained. (DCOFk1 can drop an incoming flit if the private buffers are full.) In our analysis we assume DCOFk1 uses a small shared receive buffer, connected to the N-1 private receive buffers.

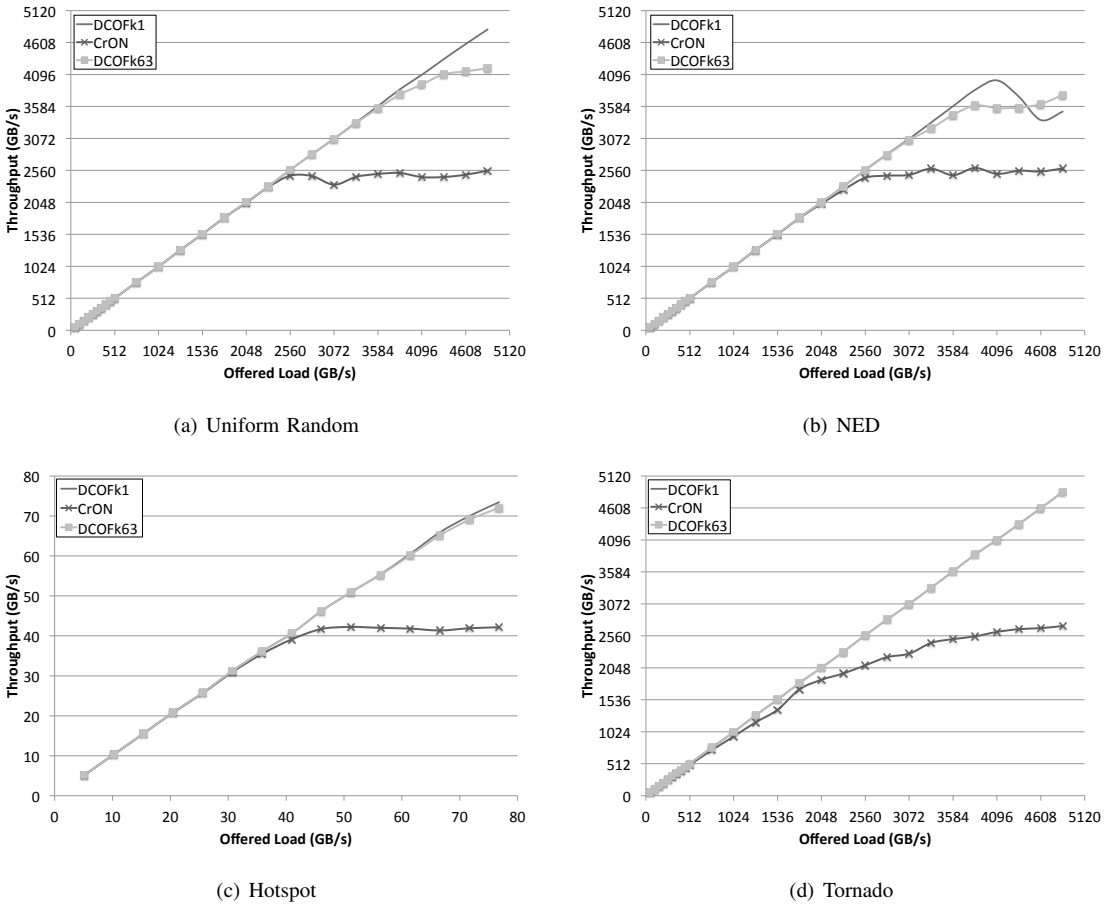
In CrON we assume each node has a shared receive buffer, since there is only one receiver per node. The amount of buffering must match the token size, so in order to avoid wasting photonic power the receive buffer size was chosen to be 16 flits (which evenly divides into the 64 wavelengths, and is also the assumption in [32]). DCOFk1 does not require a private buffer for each transmitter, only one per k (since only k simultaneous transmissions are possible). We assume a single shared transmit buffer for DCOFk1, and the shared buffer was chosen to be 32 flits since it works well with the ARQ scheme. The small shared receive buffer also stores 32 flits, to match the size of the transmit buffer.

In order to determine the optimal amount of buffering for CrON and DCOFk1, the throughput of the networks with various buffering configurations was compared to that of an equivalent network with infinitely large buffers. The NED traffic pattern was used because its behavior is similar to real traces. The results of the buffering analysis showed that CrON had degraded throughput when only 4 flit buffers were employed, and had no loss in throughput when 8 flit buffers per transmitter were available. The performance of DCOFk1 was diminished when only 2 flit buffers were used (even assuming a 2-output port local crossbar), but using a 4 flit buffers per receiver resulted in maximal throughput for the topology. Thus, the performance and power results presented in the remainder of this paper assume 8 flit buffers per transmitter and 16 flit buffers per receiver for CrON, and 32 flit transmit buffers, 4 flit receive buffers and a 32 flit shared receive buffer for DCOFk1. This results in a total of 520 and 316 flit buffers per node for CrON and DCOFk1, respectively.

B. Performance Results

The synthetic traffic "traces" provided an average offered load with an average packet size of 4 flits per packet, using a burst/lull distribution. The burst/lull injection distribution was chosen over a Bernoulli distribution since real traffic tends to be more "bursty" in nature. The throughput in GB/s is shown as a function of offered load in GB/s for DCOFk1, DCOFk63 and CrON in Figure 4. DCOFk1 and DCOFk63 outperform CrON on every one of the synthetic traffic patterns. Note that for the *hotspot* traffic pattern the offered load is limited to 80GB/s, since the maximum throughput of a single node is 80GB/s and any offered load above that is guaranteed to overwhelm any network, regardless of topology. Note also that the throughput for DCOFk1 with the NED traffic pattern does not maintain a maximum level, but actually tapers off as a higher load is offered. This is due to the ARQ flow control - as the offered load increases, more flits are dropped and must be retransmitted.

The reader may notice that the performance of DCOFk1 and DCOFk63 on the uniform random and NED traffic patterns (seen in Figure 4(a) and 4(b)) seem incorrect, since DCOFk1 has a higher throughput than DCOFk63 under high loads – this result is due to the way the flow signal is re-enabled in DCOFk63. DCOFk63 could perform as well as DCOFk1 if the same ARQ-based flow control scheme were employed, or if it used a re-enabling algorithm other than round robin. In Figure 4(d) the DCOFk1 results are indistinguishable from the DCOFk63 results since both networks perform ideally whenever the traffic pattern has a single source for each possible destination.



(a) Uniform Random

(b) NED

(c) Hotspot

(d) Tornado

Fig. 4: Throughput (GB/s) vs. Offered Load (GB/s)

From the graphs it appears that DCOFk1 performs ideally on all traffic patterns except for NED. In reality, the performance of DCOFk1 is slightly lower than the ideal starting at 56GB/s for hotspot and 4096GB/s for uniform random. The performance of DCOF does match the ideal for tornado, and this would also be true for nearest neighbor, transpose, bit inverse, and any other synthetic traffic pattern where each destination can only receive from a single source. This holds because DCOF does not require arbitration in order to send a flit, so it is not possible for a single source to trigger the need to drop a flit.

The average flit or packet latency is another common metric which is used to compare networks. We decided to look in more detail at the *components* of the average flit latency. Figure 5 shows the average flit latency component due to arbitration in CrON and flow control in DCOFk1 and DCOFk63 when using the NED traffic pattern.[4] Note that arbitration in CrON adds latency to each flit even under low loads, but the flow control in both DCOFk1 and DCOFk63 only adds latency when the network has become overwhelmed. As was stated earlier, arbitration is an overhead that must be paid whenever communicating, while both the ARQ and enable/disable flow control is an ”on-demand” penalty that is only paid when the network is overwhelmed.

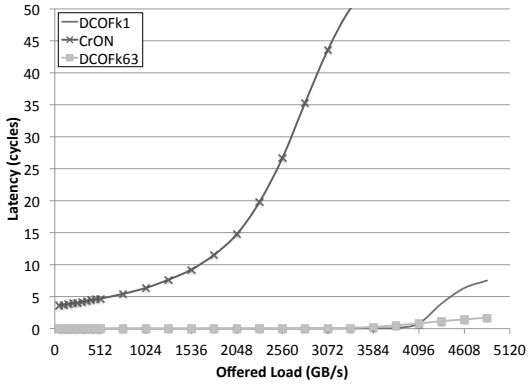
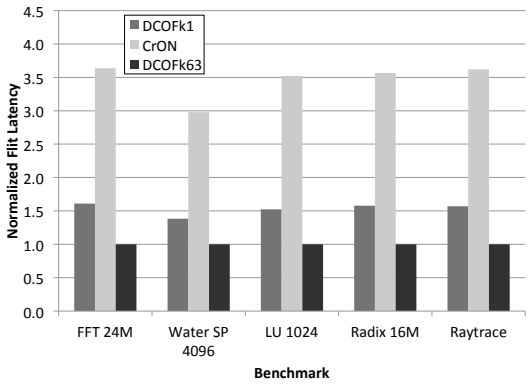


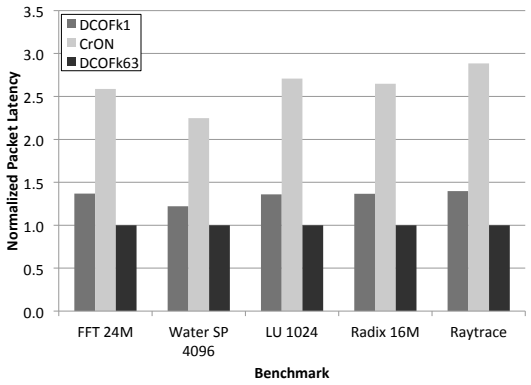
Fig. 5: Latency (in cycles) vs. Offered Load (GB/s) for the NED Traffic Pattern

The performance results of the SPLASH-2 runs are shown in Figure 6. Figures 6(a) and 6(b) show the average flit and packet latencies for DCOFk1, DCOFk63 and CrON, normalized to the network with the lowest latency (in all cases DCOFk63). The figures show that DCOFk1 and DCOFk63 have dramatically lower average latencies across all the benchmarks; however, the lower latency does not result in as dramatic a difference in the overall execution time.

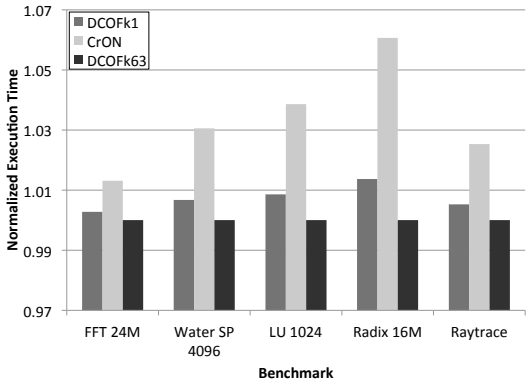
[4]NED was chosen because the flow control component in DCOF is by far the highest in NED - it is negligible in the other traffic patterns.



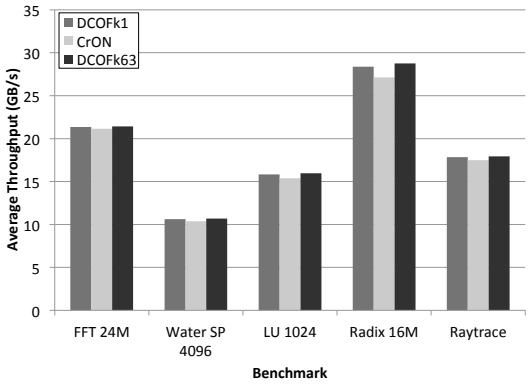
(a) Flit Latency, Normalized to DCOFk63



(b) Packet Latency, Normalized to DCOFk63



(c) Execution Time, Normalized to DCOFk63



(d) Average Throughput in GB/s

Fig. 6: SPLASH-2 Performance Results

Figure 6(c) shows the execution time of each benchmark normalized to the shortest execution time, and the figure shows that DCOFk63 executed the benchmarks from 1.3% to 6% faster than CrON and from 0.3% to 1.3% faster than DCOFk1. The reader may be left wondering why reducing the packet latency by over a factor of 2 would result in such a small decrease in execution time; the answer is that the average required network throughput for the benchmarks is quite low when compared to the networks capabilities.

Figure 6(d) shows the average throughput in GB/s for the various benchmarks. The average throughput of the SPLASH-2 benchmarks equates to ∼0.4% of the total network bandwidth for DCOFk1 and CrON, leading one to question the wisdom of building a network like this. While it may at first appear that the networks are over-designed, it is important to note that the average of the *peak* throughputs attained on the benchmarks was ∼25.3% of the total available network bandwidth for CrON, and ∼99.7% for DCOFk1. In other words, at some point while executing on DCOFk1 the maximum possible network throughput was obtained on every benchmark except for Radix, indicating there are critical points at which all the network bandwidth is utilized. More importantly, however, is the fact that one must be extremely careful not to unwisely restrict the flexibility of tomorrow's on-chip processor network based on the results of running yesterday's parallel processing benchmarks.

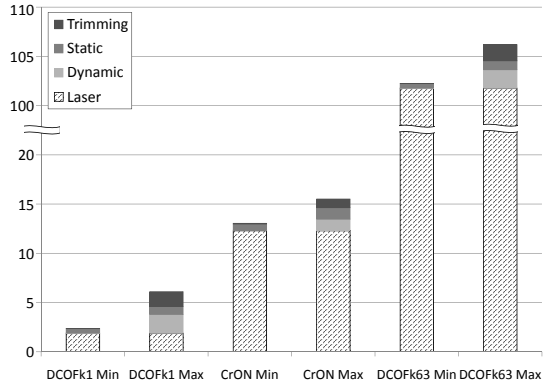


Fig. 7: Power (W) vs. Network (Min/Max Load). (Note that there is a discontinuity in the scale between 20 and 100)

C. Power Results

The minimum and maximum power consumption for DCOFk1, DCOFk63 and CrON is shown in Figure 7. The minimum power consumption is the minimum power that must be consumed even when the network is idle and at its lowest ambient temperature, while the maximum power is the

maximum observed across all the simulations. The dominant factor for all three networks is the laser power, which is consumed regardless of activity. The reader may notice that CrON also consumes dynamic electrical power even when idle; this is due to the fact that arbitration tokens must be replenished every loop, requiring modulation of the arbitration microrings. It should also be clear that the amount of photonic power required for DCOFk63 is substantial, and will likely limit the number of designs that can employ DCOFk63.

As one might expect, the overall maximum trimming power required for DCOFk1 is higher than for CrON, since DCOFk1 has ∼88% more microrings. However, the average trimming power *per microring* is actually 18% higher for CrON. We observed in [28] that the heating power required for trimming has a non-linear relationship with microring count, and our findings show that current injection has a non-linear relationship as well. CrON requires more trimming power per microring since the network operates at a higher temperature due to the greater power consumption when compared to DCOFk1.

The maximum amount of dynamic power consumed by DCOFk1 is much higher than that of CrON, but DCOFk1 also greatly outperforms CrON in the maximal case. Figure 8(a) shows the energy efficiency in fJ/b as a function of offered load in GB/s. The energy efficiency shown in Figure 8(a) is calculated by taking the power consumed divided by the actual network throughput (not the theoretical maximum throughput). The solid lines for DCOFk1, DCOFk63 and CrON are the average energy efficiencies (the average power consumed divided by average throughput), while the dotted lines show the minimum and maximum energy efficiencies for the three networks; the actual efficiency varies with achieved throughput and ambient temperature. DCOFk1 is clearly more energy efficient than CrON - percentage wise the result is most apparent under high offered load (since CrON is unable to actually achieve higher throughputs), while the greatest absolute efficiency difference can be seen under lower loads. It is also clear that CrON is more energy efficient than DCOFk63, primarily due to the tremendous amount of unused bandwidth in DCOFk63.

In the best case DCOFk1, CrON, and DCOFk63 approach 109, 652, and 2,675 fJ/b respectively, though this only occurs under high load. The energy efficiency of DCOFk63 under a 100% workload approaches 77 fJ/b, which is over a factor of 8 lower than CrON under 100% load (although it is unclear if there would ever be a time in a real system when every node would be simultaneously transmitting to every other node).

The energy efficiencies that can be obtained by DCOFk1, DCOFk63, and CrON under high load are not observed when the networks execute the Splash benchmarks, which can be seen in Figure 8(b). The average energy efficiency for DCOFk1, CrON, and DCOFk63 on the SPLASH-2 benchmarks was 24.1, 104, and 750 thousand fJ/b, respectively. The lower energy efficiency observed in these photonic networks under low load is a problem that will likely be shared with future on-chip electrical networks; while electric networks will not have the static laser overhead, the static electrical leakage is of greater and greater concern as we move from deep submicron into nanoscale technologies.

A network with lower performance may have the potential for higher energy efficiency, but a lower performing network will also impact the energy efficiency of the cores and caches due to the increased number of stalled cycles. Examining the impact of network performance on the energy efficiency of the cores is beyond the scope of this work.

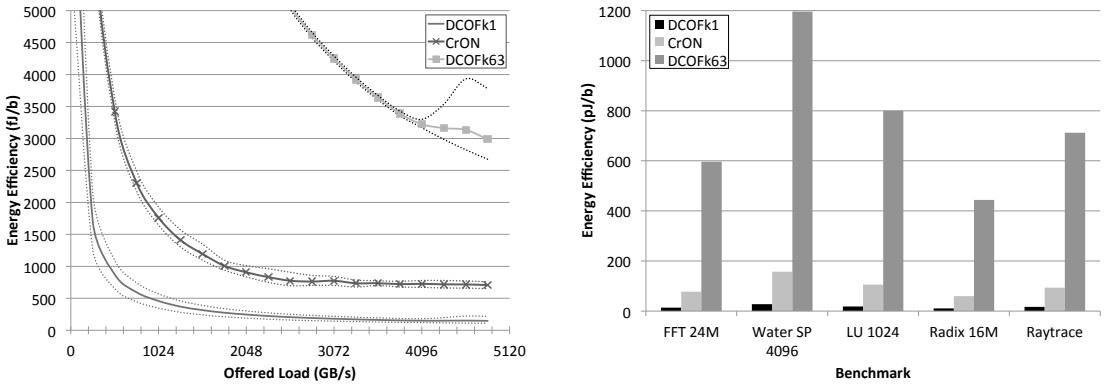
VII. Discussion

Average energy efficiency is a common concern among computer architects. As was shown in the previous section, the average throughput of the SPLASH-2 benchmarks is very low compared to the total network bandwidth, and this low average throughput leads to low average energy efficiency. However, reducing the capabilities of the network is not necessarily desirable, since the entire network bandwidth *is* utilized at certain points in the benchmarks. The main reason for the energy inefficiency at low load is the large amount of static power overhead (the static leakage and fixed laser power). Reducing the static leakage power is a well-studied area, but the approach of reducing the fixed laser power or adjusting it to match the workload has not yet been examined.

At this point scaling the laser power is not a viable option, since lowering the incoming laser energy uniformly drops the power on all links. However, it is possible the unused energy could be recaptured – the photons not used to communicate could be captured and turned into electricity. Converting the unused photons to electrons would be relatively straightforward, requiring only the modification of existing photodiode structures. The number of photons available for recapture is a function of the activity occurring on each wavelength, which is related to the workload and the distribution of ones and zeros. The theoretical limits of photonic energy recapture efficiency can be established using thermodynamic arguments, in the same way it was done for solar cells in [42]. The results indicate a peak efficiency of ∼79%, which provides a definite theoretical bound for recapture efficiency.

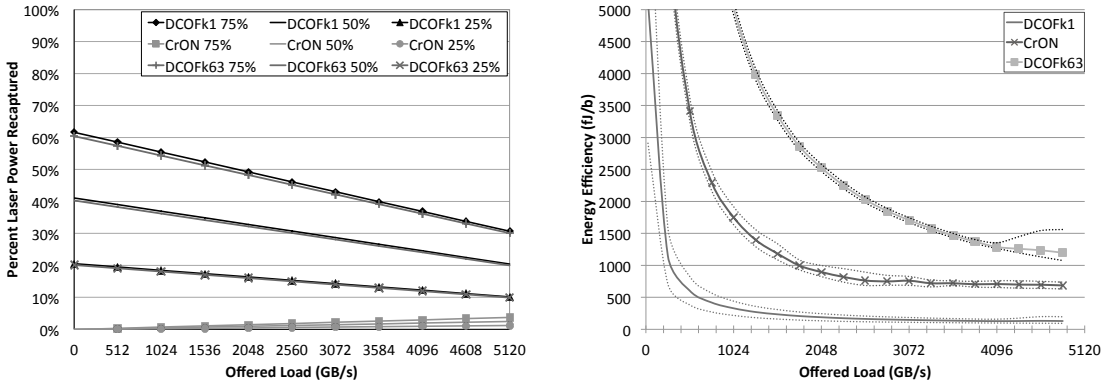
Figure 9(a) shows the percent of laser power that potentially could be recaptured for DCOFk1, DCOFk63 and CrON assuming a 25%, 50% and 75% conversion efficiency. Notice that the slope of lines for DCOF and CrON are opposite - DCOF can recapture the most photonic power under low load, while CrON recaptures the most under high load. This fact is due to the structure of the two networks; CrON can only recapture a wavelength when a zero is being sent on that bit, while DCOF can recapture a wavelength whenever a one is not being sent on that bit (recapture occurs when there is no transmit or a transmit of a zero). Another reason DCOF has a higher recapture percentage is that recapture always occurs upstream, where CrON would recapture potentially anywhere along the serpentine. These projections show that photonic recapture has the potential for substantially improving the energy efficiency of DCOF under low load, but the energy efficiency of CrON will only improve under high load (which unfortunately is the opposite of what is desired).

Figure 9(b) shows the energy efficiency in fJ/b of DCOFk1, CrON, and DCOFk63 vs. offered load in GB/s with 75%



(a) Offered Load (GB/s) (b) SPLASH-2 Benchmark

Fig. 8: Energy Efficiency in (fJ/b) vs. Offered Load (GB/s) (a) and in (pJ/b) vs. SPLASH-2 Benchmark (b)



(a) Percent of Laser Power Recaptured (b) Energy Efficiency in (fJ/b)

Fig. 9: Percent of Laser Power Recaptured (a) and Energy Efficiency in (fJ/b) (b) vs. Offered Load (GB/s)

efficient recapture. Comparing this figure to Figure 8(a), we see that the results are almost identical for CrON, but the energy efficiency for DCOFk1 and DCOFk63 is noticeably improved, especially for low offered load. Comparing the results with Figure 8(a) it is clear that recapture has the greatest impact at low load for DCOF. What is not discernible from the figures is that the idle power for DCOFk1 is cut almost in half in the best case, when using 75% efficient recapture.

Another common concern of architects is the scalability of network topologies. A 64-bit DCOFk1 with 128 nodes will require an area of $\sim$293mm^2, but a 256 node DCOFk1 would require $\sim$1,650mm^2. The photonic power of DCOFk1 does not scale linearly, although there is a less than 5% increase in required channel power scaling from 64 to 128 nodes. A 64-bit CrON with 256 nodes will require a smaller area ($\sim$323mm^2), but the photonic power of CrON will likely not scale to even 128 nodes. The number of off-resonance rings which light must pass through will roughly double when scaling CrON from 64 to 128 nodes, and this fact alone will in turn increase the path attenuation by over 6dB. Our estimates show that a 128 node CrON would require over 100W of photonic power - thus, while the scalability of DCOFk1 is limited to 128 nodes, CrON is limited to half that.

The bandwidth capability of DCOFk1 is likely sufficient to support multiple cores per node. As was shown by the SPLASH-2 benchmark performance results, the average network utilization is quite low. It is probable that a network using DCOFk1 would electrically cluster multiple cores which would become a node, as was assumed in [11]. The number of clustered cores which DCOFk1 could support could also be increased by increasing the degree of simultaneous communication k. Examining the potential for clustering cores in DCOFk1 is something we will explore in the future. In addition, we show in [24] that it is possible to create a hierarchy of optical networks in order to scale the number of processors even further. And according to Keckler [43], while it will be possible to place hundreds of processing nodes on a chip in the future, it is highly unlikely we will be able to power and/or cool them all. So scaling to 256 nodes will almost certainly be sufficient.

VIII. Related Work

Within the research community there has been a growing interest in harnessing the benefits of optics in order to address the shortcomings of electrical interconnects. In [11] HP researchers describe a 64x64 WDM based crossbar (called

Corona) for a 256-core CMP. Corona uses a multiple-writer single reader crossbar architecture, which requires arbitration (realized using a distributed scheme and additional optical channels). Cornell researchers described a bus-based scheme to connect clusters of processors in [12], and more recently propose a hybrid opto-electronic on-chip network called Phastlane that uses a low complexity nanophotonic crossbar supported by an electrical network for buffering and arbitration. Phastlane uses packets with a single flit and an ARQ based flow control scheme, where packets are allowed to be dropped. DCOFk1 uses a similar flow control scheme, with the exception that it is ACK instead of NAK based.

MIT and Berkley researchers [19] propose a multistage Clos network using a mixture of electronic routers that are connected by WDM based photonic links. Clearly, this network has less flexibility and a higher average hop-count than a crossbar. Furthermore, the CMXBar described in the paper requires arbitration, which DCOF does not. The authors in [15] propose a photonic 2D torus network that employs an electrical network for arbitration and flow control. The network is evaluated on a variety of synthetic and scientific benchmarks [20] to show that the hybrid photonic torus network can achieve a factor of 37x improvement in performance per energy spent. This paper also points out that many scientific workloads exhibit communication patterns that change over time, another reason the directly connected nature of DCOF is so attractive.

Firefly [14] is another hybrid opto-electronic network proposal that uses an electrical network for intra-cluster communication and a nanophotonic crossbar for inter-cluster communication. The Single Writer Multiple Reader (SWMR) network discussed in [14] requires a broadcast network in order to send the head flit, and this broadcast network will require arbitration - the timing between the sending of the head flit and transmitting the data flits will also require precise delay. In addition, the broadcast network will require power, which is likely to be nearly equal to that of the SWMR crossbar itself.

The FexiShare network is a flexible photonic crossbar [44] that is a combination of a Multiple Write Single Read (MWSR) and a SWMR design. The FlexiShare network decouples the number of communication channels from the number of number of nodes, in an attempt to reduce the required photonic power. FlexiShare implements a token stream for arbitration and credit sharing, adopting the reservation assisted scheme from Firefly. Recently the authors of [44] proposed an optical arbitration scheme that includes Quality of Service called FeatherWeight [45].

Sun Labs/Oracle researchers [46] recently investigated using silicon photonics for the interconnection network of a multi-chip system or "Macrochip". They analyzed three different photonic networks in the multi-die system that used mirrors to couple light between dies, and concluded that a statically routed point-to-point network outperformed the other networks analyzed. The point-to-point networks analyzed in [46] were limited to 2-bit site-to-site connections, which the authors admit "is a potential performance limiter". The inter-layer coupler assumed in [46] differs from our photonic vias in that the coupler connects signals between two dies, while our photonic via couples between layers of the same die.

IX. Conclusions

In this paper we have shown that by using multiple photonic layers, it is possible to provide a family of arbitration free topologies that are not realizable using conventional electronics. We accomplish this by creating a directly connected *fabric* of waveguides that can be configured to support everything from a crossbar to fully connected topologies. The advantages of directly (fully) connected topologies are well known - they offer the highest bisection bandwidth and are far more resilient to failures on links, since packets can be routed through unaffected nodes. Perhaps more importantly, they make writing programs easier, particularly when there are a large number of processors involved - the programmer does not have to worry about data location, for example, and anything that can be done to make parallel programming simpler is of great value to the entire community.

We have presented a detailed description of how the fabric can support two representative network instantiations, which we call DCOFk1 and DCOFk63, as well as an in-depth analysis of their performance and power consumption when compared to an optical crossbar (CrON) based on Corona [3]. We have shown the advantages of flow control over arbitration - arbitration is an overhead that is incurred whether or not it is needed, while flow control is a penalty paid only when the network is overwhelmed. This fact contributed to our observation that even though DCOFk1 and CrON have identical link, bi-sectional and total bandwidth in theory, DCOFk1 out performs CrON while consuming less power.

We found that the energy efficiency of all networks under low load is dramatically lower than it is under high load, potentially leading one to consider designing a lower performing network; however, DCOFk1 reached maximum total throughput on all but one of the SPLASH-2 benchmarks, meaning that there are certain points at which all the network bandwidth is utilized. More importantly, one must keep in mind that the SPLASH-2 benchmarks are old, and one has to be very careful when designing tomorrow's machine using yesterday's programs. Fully connected topologies offer much more flexibility and resilience, and can easily adapt to future changes in workloads. It would be far wiser to instead work towards making the energy efficiency consistent regardless of load - something photonic energy recapture has the potential to do.

References

[1] S. H. Fuller and L. I. Millett, *The Future of Computing Performance: Game Over or Next Level*. Washington, D.C.: The National Academies Press, 2011. 1

[2] D. Miller, "Rationale and challenges for optical interconnects to electronic chips," *Proceedings of the IEEE*, vol. 88, no. 6, pp. 728–749, Jun 2000. 1

[3] J. Ahn, M. Fiorentino *et al.*, "Devices and architectures for photonic chip-scale integration," *Applied Physics A: Materials Science & Processing*, vol. 95, pp. 989–997, Jun. 2009. 1, 3, 7, 13

[4] D. Miller, "Device requirements for optical interconnects to silicon chips," *Proceedings of the IEEE*, vol. 97, no. 7, pp. 1166 –1185, july 2009. 1

[5] R. Beausoleil, P. Kuekes *et al.*, "Nanoelectronic and nanophotonic interconnect," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 230–247, Feb. 2008. 1

[6] L. Chen, N. Sherwood-Droz, and M. Lipson, "Compact bandwidth-tunable microring resonators," *Opt. Lett.*, vol. 32, no. 22, pp. 3361–3363, 2007. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-32-22-3361 1

[7] Q. Xu, B. Schmidt *et al.*, "Micrometre-scale silicon electro-optic modulator," *Nature*, vol. 435, no. 7040, pp. 325–327, 2005. [Online]. Available: http://dx.doi.org/10.1038/nature03569 1

[8] L. Zhou, S. Djordevic *et al.*, "Design and evaluation of an arbitration free passive optical crossbar for on-chip interconnection networks," *Applied Physics A - Materials Science Engineering - Special Issue on Photonics Interconnects*, vol. 95, no. 4, pp. 1111–1118, June 2009. 1

[9] B. R. Koch, A. W. Fang *et al.*, "Mode-locked silicon evanescent lasers," *Optical Express*, vol. 15, no. 18, pp. 11 225–11 233, 2007. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-18-11225 1

[10] M. Lipson, "High performance photonics on silicon," *Optical Fiber Communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, pp. 1–3, Feb. 2008. 1

[11] D. Vantrease, R. Schreiber *et al.*, "Corona: System implications of emerging nanophotonic technology," in *ISCA '08: Proceedings of the 35th International Symposium on Computer Architecture*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 153–164. 1, 5, 12

[12] N. Kirman, M. Kirman *et al.*, "Leveraging optical technology in future bus-based chip multiprocessors," in *MICRO 39: Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 492–503. 1, 13

[13] M. J. Cianchetti, J. C. Kerekes, and D. H. Albonesi, "Phastlane: a rapid transit optical routing network," *SIGARCH Comput. Archit. News*, vol. 37, no. 3, pp. 441–450, 2009. 1

[14] Y. Pan, P. Kumar *et al.*, "Firefly: illuminating future network-on-chip with nanophotonics," *SIGARCH Comput. Archit. News*, vol. 37, no. 3, pp. 429–440, 2009. 1, 13

[15] A. Shacham, K. Bergman, and L. P. Carloni, "On the design of a photonic network-on-chip," in *NOCS '07: Proceedings of the First International Symposium on Networks-on-Chip*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 53–64. 1, 13

[16] ——, "The case for low-power photonic networks on chip," in *DAC '07: Proceedings of the 44th annual Design Automation Conference*. New York, NY, USA: ACM, 2007, pp. 132–135. 1

[17] A. Shacham and K. Bergman, "Building ultralow-latency interconnection networks using photonic integration," *IEEE Micro*, vol. 27, no. 4, pp. 6–20, 2007. 1

[18] C. Batten, A. Joshi *et al.*, "Building manycore processor-to-dram networks with monolithic silicon photonics," in *HOTI '08: Proceedings of the 2008 16th IEEE Symposium on High Performance Interconnects*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 21–30. 1

[19] A. Joshi, C. Batten *et al.*, "Silicon-photonic clos networks for global on-chip communication," in *NOCS '09: Proceedings of the 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 124–133. 1, 13

[20] G. Hendry, S. Kamil *et al.*, "Analysis of photonic networks for a chip multiprocessor using scientific applications," *Networks-on-Chip, International Symposium on*, vol. 0, pp. 104–113, 2009. 1, 13

[21] J. Shalf, S. Kamil *et al.*, "Analyzing ultra-scale application communication requirements for a reconfigurable hybrid interconnect," in *SC '05: Proceedings of the 2005 ACM/IEEE conference on Supercomputing*. Washington, DC, USA: IEEE Computer Society, 2005, p. 17. 1

[22] G. Ballard, J. Demmel, and A. Gearhart, "Communication bounds for heterogeneous architectures," University of California, Berkeley, Berkeley, CA, USA, Tech. Rep., 2011. [Online]. Available: http://digitalassets.lib.berkeley.edu/techreports/ucb/text/EECS-2011-13.pdf 1

[23] L. Zhou, K. Kashiwagi *et al.*, "Towards athermal optically-interconnected computing system using slotted silicon microring resonators and rf-photonic comb generation," *Applied Physics A: Materials Science & Processing*, vol. 95, pp. 1101–1109, 2009, 10.1007/s00339-009-5120-7. [Online]. Available: http://dx.doi.org/10.1007/s00339-009-5120-7 2

[24] C. J. Nitta, "Design and analysis of large scale nanophotonic on-chip networks," Ph.D. dissertation, University of California, Davis, 2011. 3, 4, 5, 6, 12

[25] D. Taillaert, P. Bienstman, and R. Baets, "Compact efficient broadband grating coupler for silicon-on-insulator waveguides," *Opt. Lett.*, vol. 29, no. 23, pp. 2749–2751, 2004. [Online]. Available: http://ol.osa.org/abstract.cfm?URI=ol-29-23-2749 3

[26] G. Maire, L. Vivien *et al.*, "High efficiency silicon nitride surface grating couplers," *Opt. Express*, vol. 16, no. 1, pp. 328–333, 2008. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-16-1-328 3

[27] J. Dionne, L. Sweatlock *et al.*, "Silicon-based plasmonics for on-chip photonics," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16, no. 1, pp. 295 –306, jan.-feb. 2010. 3

[28] C. Nitta, M. Farrens, and V. Akella, "Addressing system-level trimming issues in on-chip nanophotonic networks," in *High Performance Computer Architecture, 2011. HPCA 2011. IEEE 17th International Symposium on*, Feb. 2011. 3, 11

[29] L. Zhou, K. Okamoto, and S. Yoo, "Athermalizing and trimming of slotted silicon microring resonators with uv-sensitive pmma upper-cladding," *Photonics Technology Letters, IEEE*, vol. 21, no. 17, pp. 1175–1177, Sept.1, 2009. 3

[30] C. Nitta, M. Farrens, and V. Akella, "Resilient microring resonator based photonic networks," in *Micro-44: Proceedings of the 44nd Annual IEEE/ACM International Symposium on Microarchitecture*, Dec. 2011. 3

[31] ——, "Dcaf - a directly connected arbitration-free photonic crossbar for energy-efficient high performance computing," in *Parallel Distributed Processing (IPDPS), 2012 IEEE International Symposium on (to appear)*, May 2012. 5

[32] D. Vantrease, N. Binkert *et al.*, "Light speed arbitration and flow control for nanophotonic interconnects," in *Micro-42: Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*. New York, NY, USA: ACM, 2009, pp. 304–315. 6, 8

[33] H.-S. Wang, X. Zhu *et al.*, "Orion: A power-performance simulator for interconnection networks," *Microarchitecture, IEEE/ACM International Symposium on*, vol. 0, p. 294, 2002. 7

[34] S. Thoziyoor, N. Muralimanohar *et al.*, "Cacti 6.0: A tool to model large caches," HP Laboratories, Palo Alto, CA, USA, Tech. Rep. HPL-2009-85, Apr. 2009. [Online]. Available: http://www.hpl.hp.com/techreports/2009/HPL-2009-85.pdf 7

[35] Semiconductor Industry Association, "International technology roadmap for semiconductors 2009," Semiconductor Industry Association, Tech. Rep., 2009. [Online]. Available: http://public.itrs.net/Links/2009ITRS/Home2009.htm 7

[36] R. Ho, "On-chip wires: Scaling and efficiency," Ph.D. dissertation, Stanford University, 2003. 7

[37] A. Naeemi, J. Xu *et al.*, "Optical and electrical interconnect partition length based on chip-to-chip bandwidth maximization," *Photonics Technology Letters, IEEE*, vol. 16, no. 4, pp. 1221 –1223, Apr. 2004. 7

[38] W. Huang, K. Sankaranarayanan *et al.*, "Accurate, pre-rtl temperature-aware design using a parameterized, geometric thermal model," *IEEE Transactions on Computers*, vol. 57, no. 9, pp. 1277–1288, 2008. 7

[39] D. Vantrease and N. Binkert, personal communication about photonic token design in Corona, 2011. 7

[40] C. Nitta, K. Macdonald *et al.*, "Inferring packet dependencies to improve trace based simulation of on-chip networks," in *Networks-on-Chip (NOCS), 2011 Fifth ACM/IEEE International Symposium on (to appear)*, May 2011. 8

[41] A.-M. Rahmani, I. Kamali *et al.*, "Negative exponential distribution traffic pattern for power/performance analysis of network on chips," in *VLSID '09: Proceedings of the 2009 22nd International Conference on VLSI Design*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 157–162. 8

[42] J. Nelson, *The Physics of Solar Cells*. London: Imperial College Press, 2003. 11

[43] S. Keckler, "Life after dennard and how i learned to love the picojoule," in *Micro-44: Proceedings of the 44nd Annual IEEE/ACM International Symposium on Microarchitecture*, Dec. 2011, p. ii. 12

[44] Y. Pan, J. Kim, and G. Memik, "Flexishare: Channel sharing for an energy-efficient nanophotonic crossbar," in *High Performance Computer Architecture, 2010. HPCA 2010. IEEE 16th International Symposium on*, Jan. 2010. 13

[45] ——, "Featherweight: low-cost optical arbitration with qos support," in *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO-44 '11. New York, NY, USA: ACM, 2011, pp. 105–116. [Online]. Available: http://doi.acm.org/10.1145/2155620.2155633 13

[46] P. Koka, M. O. McCracken *et al.*, "Silicon-photonic network architectures for scalable, power-efficient multi-chip systems," in *Proceedings of the 37th annual international symposium on Computer architecture*, ser. ISCA '10. New York, NY, USA: ACM, 2010, pp. 117–128. [Online]. Available: http://doi.acm.org/10.1145/1815961.1815977 13