Opportunities and Challenges of using Plasmonic Components in Nanophotonic Architectures

Hassan M. G. Wassel, *Student Member, IEEE,* Daoxin Dai, *Member, IEEE,* Mohit Tiwari, Jonathan K. Valamehr, *Student Member, IEEE,* Luke Theogarajan, *Member, IEEE,* Jennifer Dionne, Frederic T. Chong, *Member, IEEE,* Timothy Sherwood, *Member, IEEE*

***Abstract*—Nanophotonic architectures have recently been proposed as a path to providing low latency, high bandwidth network-on-chips. These proposals have primarily been based on micro-ring resonator modulators which, while capable of operating at tremendous speed, are known to have both a high manufacturing induced variability and a high degree of temperature dependence. The most common solution to these two problems is to introduce small heaters to control the temperature of the ring directly, which can significantly reduce overall power efficiency. In this paper, we introduce plasmonics as a complementary technology. While plasmonic devices have several important advantages, they come with their own new set of restrictions, including propagation loss and lack of Wave Division Multiplexing (WDM) support. To overcome these challenges we propose a new hybrid photonic/plasmonic channel that can support WDM through the use of photonic micro-ring resonators as variation tolerant passive filters. Our aim is to exploit the best of both technologies: wave-guiding of photonics, and modulating using plasmonics. This channel provides moderate bandwidth with distance independent power consumption and a higher degree of temperature and process variation tolerance. We describe the state of plasmonics research, present architecturally-useful models of many of the most important devices, explore new ways in which the limitations of the technology can most readily be minimized, and quantify the applicability of these novel hybrid schemes across a variety of interconnect strategies. Our link-level analysis show that the hybrid channel can save from 28% to 45% of total channel energy-cost per bit depending on process variation conditions.**

***Index Terms*—nanophotonics, plasmonics, on-chip interconnects, energy-efficient network-on-chip**

I. Introduction

As we continue to add more cores, accelerators, and other communication hungry resources on die, delivering low-power high-bandwidth interconnect becomes increasingly critical. These demands have driven computer architects to consider new interconnect technologies not traditionally used for on-chip communication, such as optics. Optical interconnects offer high frequency operation, limited attenuation over long distances, the ability to perform wavelength division multiplexing, and low power modulation and detection. Recent advances in nanophotonics have been able to bring the structures required for non-trivial topologies down to a scale that they could feasibly be coupled with a microprocessor die.

Most of these architectural proposals have been based on the micro-ring resonator-based modulators and filters which have many attractive properties such as compactness, energy-efficiency, and the ability to tune the wavelength of resonance. However, the resonant wavelength of these rings is highly temperature dependent. Typically this problem is solved by adding a tiny control circuit containing a heater connected to the ring, thus tuning the ring temperature. In addition to temperature induced variation, the manufacturing process itself introduces variation as perturbations in the geometry of ring can change its resonant frequency [49]. Counteracting this effect can be hard to get right because it may require corrections to two adjacent rings which require two different temperatures. Moreover, heat that is generated according to the running workload can change the die temperature by 55°C [31]. Even if perfect control was achieved, trimming requires a heater that consumes around 50μW/°C [11]. This nearly static heating requirement may reduce the energy-proportionality of these photonic proposals. However, micro-ring modulators are not the only way to build a network with photonic waveguides.

Recent advances in physics and materials have demonstrated that surface plasmon polaritons (SPP), which propagate through a form of hybrid electrical/optical propagation, have several interesting properties that lend themselves to this problem. A plasmon is a quasi-particle formed from the coupling of a photon and a traveling electron density wave at the interface between a metal and a dielectric. Because these waves oscillate at optical frequencies they maintain many of the energy and bandwidth benefits of optical interconnects. However, because surface plasmon polaritons (SPP) are primarily longitudinal waves formed by charge compression, they do not require waveguides that are at the same scale as the wavelength of light. This in turn means that the waves can propagate in a fraction of the area, that there are fewer restrictions on waveguide geometry, and that modulation is in theory much easier. Unfortunately it also means, even with the best-known methods for generating and propagating plasmons, the waves are more susceptible to damping, often after only a few tens of micrometers. We can exploit these principles to build energy-efficient high-speed plasmonic modulators with much higher temperature tolerance due to its low-Q (around 100).

H. M. G. Wassel, F. T. Chong and T. Sherwood are with the Department of Computer Science at UC Santa Barbara, Santa Barbara, CA 93105 USA, email: {hwassel, chong, sherwood}@cs.ucsb.edu

D. Dai is with the Centre for Optical and Electromagnetic Research, State Key Laboratory for Modern Optical Instrumentation, Zhejiang Provincial Key Laboratory for Sensing Technologies, Zhejiang University, Zijingang Campus, Hangzhou 310058, China. email: dxdai@zju.edu.cn

M. Tiwari is with the Department of Computer Science at UC Berkeley, Berkeley, CA 94720 USA, email: tiwari@eecs.berkeley.edu

J. K. Valamehr and L. Theogarajan are with the Department of Electrical and Computer Engineering at UC Santa Barbara, Santa Barbara, CA 93105 USA, email: {valamehr, ltheogar}@ece.ucsb.edu

J. Dionne is with the Department of Material Science and Engineering at Stanford University, Stanford, CA 94305 USA, email: jdionne@stanford.edu

In this paper, we propose a new hybrid photonic/plasmonic channel that uses a photonic waveguide and plasmonic modulator. While these links have several advantages, they do not naturally support wavelength division multiplexing. Thus, we further propose the use of *low-Q* passive micro-ring filters to support wavelength division multiplexing. Contrary to typical optical system design where low-Q means poor performance, when used in conjunction with plasmonic modulators, we demonstrate that the temperature and variation tolerance of the low-Q ring filters can more than make up for the lack of selectivity in terms of energy per bit in this new design. This further allows our system to operate more efficiently than electrical-only interconnect at utilization levels more than twice as low as photonics-only networks. Of course nothing comes for free and these hybrid plasmonic/photonic networks do limit the types of designs that are possible. We examine the architectural implications of such a design both qualitatively and quantitatively. By examining several channel uses, we discovered that even with the necessary increases to the laser power due to the increased loss of plasmonic devices, using our new hybrid plasmonic/photonic structures in a point-to-point manner can save up to 45% of energy cost per bit under certain conditions of variation.

The rest of this paper is organized as follows: the next section provides a survey of the architectural proposals and a motivation for our study. Section III presents plasmonics technology and its main components. Then, Section IV describes the design of our novel hybrid plasmonic/photonic channel and how it can tolerate temperature and process variation. Then, Section V provides a link level analysis on the conditions under which the hybrid channel can save power. Section VI reports on our study on the architectural implications of such channel on optical buses, meshes and point-to-point channels. We then evaluate the effect of the channel in a network-on-chip setting in Section VII. We provide the conclusions and future work in Section VIII.

This paper has the following contributions:

1) Introducing the plasmonics as a complementary technology to photonics to the architecture community.
2) Proposing a novel hybrid photonic/plasmonic channel that is tolerant to changes due to change in temperature and process variation and quantify the conditions in which the hybrid channel will save tuning power.
3) Studying the architectural implications of this hybrid channel.

II. Related Work

In the multi and many core era, communication is a crucial aspect of system performance. While short local wires are used to connect logic gates into functional units, and long global lines distribute the clock and power, in electrical network-on-chips (NoC), the bulk of the core-to-core communication is handled by intermediate/global length wires. Simply driving these wires can account for almost half of NoC interconnect power consumption [2].

Recently, nanophotonic interconnects have been proposed as a replacement of intermediate core-to-core wires because

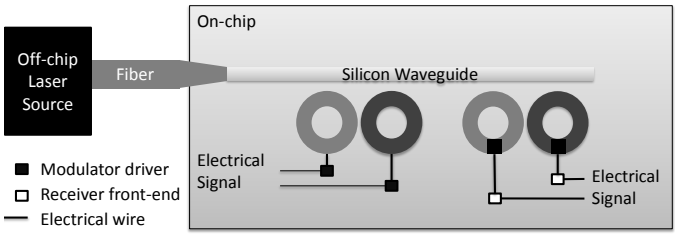


Fig. 1. A top-down diagram of a photonic link consisting of a laser source, modulator and photo-detector driven by electrical components. Laser power of varying wavelengths are brought onto the chip and routed through waveguides. Modulators and detectors are built as ring resonators tuned to different wavelengths.

of their almost distance-independent power consumption, low-latency, and high bandwidth. Figure 1 shows the main components of a photonic link, including: a ring modulator, germanium photo-detectors, and waveguides. One of the first proposed architectures was Corona [43], a fully-optical crossbar architecture using token-ring arbitration. Their later work improved the flow control algorithm by introducing a token-slot arbitration [42]. The main limitation of Corona is the huge number of ring resonators (around one million) required for modulation and detection of light which consume considerable area and power. Joshi et al. proposed an alternative technique making use of and optical clos network with electrical routers. An optical bus is used as the wiring network, connecting the electrical routers in a point-to-point fashion with almost uniform latency [20].

Firefly [37] was proposed as a hybrid optical/electrical architecture that uses an optical bus in order to provide fast links for long-distance links in a dragonfly-like topology [23]. They used electrical signaling for short distances because of the power consumption advantages electrical connectivity provides at small scales. More recently, Pan et al. suggested sharing the optical bus by not having a dedicated channel for each node in order to reduce power consumption [36]. Phastlane was a proposal of an all-optical mesh network with electrical buffering and negative acknowledgment in case of lack of buffering resources [8].

Kırman et al. proposed a wavelength-based oblivious routing for a logical torus mapped into a physical optical bus [25]. While reducing the power consumption considerably because of the predetermined path, the latency per packet may increase as packets wait for wavelengths that might be in use by another node to communicate to the same destination node. Another application of nanophotonics is trying to replace the DRAM bus with an optical bus [3], [4].

While the above architectures have demonstrated some of the potential of photonic interconnect, the technology certainly has its limitations, including: diffraction-limited sizes, temperature dependence, and manufacturing variability. First, nanophotonic components are governed by the diffraction limit which dictates that light cannot be confined in a space smaller than approximately $\lambda/2n$ where λ is the light wavelength and n is the refractive index of the material. This means that all of the waveguides, modulators, and ring filters must be built at micrometer scales or larger to use the C-band of around 1550 nm wavelength. This size mismatch between

micrometer-scale photonic components and nanometer-scale electronic component limits the integration viability of both technologies on the same chip using the same process. Moreover, the large-sized components make designing interference-based modulators not viable from an on-chip integration point of view and limit on-chip modulation options to the micro-ring resonators that are temperature and variation dependent. This brings us to our second point, photonic components are temperature-dependent which can affect signal integrity when integrated in a microprocessor chip with variable temperatures. This temperature dependence does have an upside however, as it can be exploited to allow the dynamic tuning of micro-ring modulators, allowing the resonance range of those modulators to be adjusted using heating. This heating requirement is estimated to consume around 100 fJ/bit [3]. Nitta et al. proposed the usage of redundant rings to help reduce the thermal tuning power by using these redundant rings in case temperature shifts in that region in a sliding window fashion [33]. Moreover, they assume that heaters are not required per ring because rings deviate together. However, that does not take into account the thermal tuning that is done to correct process variation error that might not be physically co-located as the case in temperature variation. Their paper only considered heat generated by the optical losses but not the chip temperature change according to workload. The devices being developed in the plasmonics community offer some new opportunities, if architected in the right way, to side step some of these limitations.

III. Introduction to Plasmonics

In this section, we introduce plasmonics technology. We start by describing the surface plasmon polariton phenomenon and we then describe basic components of this technology with their limitations and opportunities.

A. Surface Plasmon Polaritons Background

Surface plasmon polaritons (SPP) are electromagnetic waves that are coupled to free electron collective oscillations in a metal. As shown in Figure 2, when a light beam impinges on a metal-dielectric interface with a certain angle, surface plasmon polaritons are excited and propagate along the surface of the metal. Interestingly, surface plasmons excited at the interface of a metal and dielectric maintain the frequency of the exciting light, while at the same time having a much shorter wavelength. This shorter wavelength allows the construction of nanoscale waveguides and devices that tightly confine even very high frequency electromagnetic waves, in a way side-stepping traditional diffraction limits. Propagation of the SPP, however, will be limited by both metal absorption (i.e. ohmic losses) and free-space radiation. Noticing the importance of this emerging field, the term "plasmonics" was coined in 2000 [1] to define the study of this phenomenon. During the last decade, the plasmonics field has made significant progress in improving the propagation distances of SPP modes and recently, has contributed considerable new work on active components and plasmonic sources. In the rest of this section, we discuss research aimed at improving propagation distances and CMOS-compatibility.

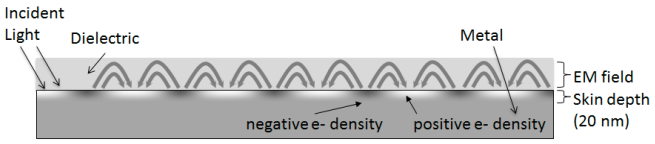


Fig. 2. Optical excitation of surface plasmon polaritons (SPP) along a metal-dielectric interface produces a longitudinal charge-density wave with a wavelength and skin depth much smaller than the illumination wavelength.

SPP modes can be excited in a variety of metal surface configurations, and one way of improving propagation distances and wave confinement is through the use of multiple metal-dielectric interfaces. There are two types of multi-interface structures: metal-insulator-metal (MIM) (also known as metal-dielectric-metal or slot waveguides) and insulator-metal-insulator (IMI) structures. In these structures, two modes can propagate, resulting from coupling of surface plasmons along each metal-dielectric interface: a magnetic field-symmetric mode and a field-anti-symmetric mode. In the IMI structures, the symmetric mode experiences less attenuation than the anti-symmetric mode, due to its lower field confinement in the metal. Because of this mode's low loss, it is often called the long-range SPP (LR-SPP) and, correspondingly, the IMI structure is called the Long-Range SPP (LR-SPP) waveguide. However, IMI waveguides suffer from poor mode confinement, with plasmon fields often penetrating many microns into the surrounding dielectric [12], [48]. Consequently, there is no system-level benefit from using them rather than conventional photonic waveguide.

For any new technology to be commercially viable, it has to be CMOS process compatible. Although silver and gold generally exhibit the lowest plasmonic losses (and thus they are the defacto materials used in plasmonics studies), copper and aluminum are also good plasmonic materials. For example, it is shown that Al and air slot waveguides provide propagation distances more than 100 microns with around a micron pitch [48]. CMOS-compatible dielectrics such as Si, SiO_2 and Si_3N_4 are also ideal for low-loss plasmon propagation [12]. In general, the longest propagation lengths are achieved for the lowest dielectric refractive index.

B. Plasmonic slot waveguides

A plasmon slot waveguide (or MDM waveguide) consists of two metal sheets separated by a thin dielectric core [14]. The thickness of the dielectric and the skin depth of the metal determine the wavelength of the light signal that can propagate in the waveguide. For maximum mode confinement, the depth of the metal should be optically opaque, as thick as the metal skin depth (approximately 20 nm). Having a metal thickness a little higher than double of the skin depth ensures that modal fields remained confined within the core and metallic cladding, leading to very low pitch of hundreds of nanometers (compared to 5.5 μm of Si waveguides). The dielectric thickness can range from just a few nanometers to over hundreds of nanometers. For a core thickness less than roughly 50 nm, only plasmonic modes will propagate. As the core thickness is increased, the waveguide begins to support

transverse electric and magnetic photonic modes, similar to microwave waveguides.

In the recent study of Si-based plasmonics [12], it is found that in terms of propagation distances, air is the best dielectric followed by SiO_2 and then Si. This result conforms to an earlier study [48] which also indicates that Al/Air is best metal/dielectric combination (although they did not compare with silver). For $\lambda = 1550$, signals can propagate up to 80 μm in a silver/SiO_2/silver waveguide with 250 nm core [14]. It is worth noting that an active area of plasmonics research is the introduction of gain material in the dielectric core of the slot waveguide to increase the propagation distance. Preliminary results show that a 27% improvement can be achieved [18].

As promising as this sounds at first, our analysis showed that plasmonic waveguides are highly unlikely to be more energy efficient than electrical signaling because electrical RC wires consume tens of fJ/bit at the scale of 100 μm. When extended to cover large propagation distances as a replacement for global electrical wires by converting the signal to the electrical domain and modulating a new CW laser every 100μm, the static laser power that needed to accommodate for the high losses of plasmonic MDM waveguides simply outweighs any other benefit (by orders of magnitude).

C. Plasmonic slot passive devices

Plasmonic slot passive devices enable efficient photonic plasmonic coupling, bending, and splitting, and are important in integrating plasmonic devices with photonic circuits. In order to integrate a hybrid photonic-plasmonic link, an efficient coupling between plasmonic and silicon waveguide must be achieved. Experiments show coupling losses as low as 1.1 dB for photonic to plasmonic mode conversion [10]. These low loss couplers are critical to our hybrid channel. Moreover, several demonstrated passive devices allow for the non-trivial routing of plasmonic circuits. Splitters, such as y-shaped and T-shaped splitters, have been shown to have only 10% losses.

D. Modulator

Several modulators have been proposed with various approaches and performance characteristics, from thermo-optical [32], all optical [35] and electro-optical [13], [7] modulator. Up to 100 femotosecond pulses have been demonstrated leading the way to THz operation. Comparison of all of these modulators was covered by MacDonald and Zheludev [29]. From this study and our assessment of [7], we conclude that the electro-optical modulators (including the PlasMOStor [13] and Compact Modulator [7]) are the most suitable for chip applications because of their fJ/bit energy consumption and high GHz operating frequency.

The PlasMOStor extends metal-oxide-silicon (MOS) technology to modulate light from an optical source to an optical drain by using the oxide as a plasmon slot waveguide. In the absence of a gate bias, both plasmonic and photonic modes can propagate through the MOS structure. By applying a gate voltage, the photonic mode experiences cutoff due to refractive index modulation of the silicon. However, the plasmon mode propagates through. Large modulation depths are achieved by engineering the separation between the source and the drain so that interference between plasmonic and photonic modes is destructive when there is no voltage applied to the gate. Consequently, the intensity at the optical drain can vary from near-zero in the off state to a finite value (equal to the plasmon intensity) in the on-state. The operating speed of the plasMOStor is limited by the speed of charge accumulation in the silicon (corresponding to refractive index modulation). So in principle, the switching speed can be as fast as a MOSFET transistor.

Modulator	Compact Modulator[7]	PlasMOStor [13]	Photonic Modulator [15]
Bandwidth (Gbps)	> 100	> 40	10
Energy per bit (fJ/bit)	1	6.8	50
Insertion loss (dB)	3	1.1	3
Modulation depth (dB)	> 3	> 10	12
Area (μm^2)	0.5	4	1000
Frequency (GHz)	> 100	59	11
Capacitance (fF)	1	14	50
Swing voltage (V)	1	0.7	2

TABLE I
MODULATOR TECHNOLOGY PARAMETERS: IT IS CLEAR THAT PLASMONIC MODULATORS HAVE VERY LOW CAPACITANCE THAT LEADS TO HIGH PERFORMANCE AND LOW POWER CONSUMPTION. HOWEVER, IT IS WORTH NOTING THAT PLASMONIC INSERTION LOSSES ARE OPTIMISTIC ESTIMATIONS IN THE RESPECTIVE PAPERS. WE WILL STUDY THE EFFECT OF INSERTION LOSS IN OUR ARCHITECTURAL ANALYSIS.

An MDM compact plasmonic modulator based on resonance in a nano slit-groove that is created in the metal layer and filled with dielectric material was proposed in [7]. Modulation is achieved by modulating the refractive index of the dielectric in the groove. The active media in the cavity can be quantum structures Stark Effect (QSSE) in a silicon-germanium system that was realized. That makes this plasmonic modulator a CMOS compatible one. It is reported that this modulator has a low-Q (around 100) cavity [7] which enables it to tolerate temperature variation [6]. It is worth noting that more experimental results are required before we can say anything about the process variation tolerance of plasmonic modulator and this paper aims to motivate such work.

As we can see from Table I, both modulators offer at least a full order of magnitude better energy per bit power consumption compared with the best-known photonic ring-based modulators [15] according to a recent survey [38]. However, this is not the fundamental limit of a photonic modulator in terms of energy per bit, with estimates that it can reach (5 fJ/bit). Moreover, the area of these modulators is at least one quarter of the area occupied by photonic modulators. When coupled with conventional a photonic waveguide using the couplers discussed above that can achieve upwards of 90% coupling efficiency [44].

IV. NOVEL HYBRID PLASMONIC/PHOTONIC CHANNEL

As we discussed in Section III-B, plasmonic waveguides either have the same pitch as photonic waveguides or suffer from huge losses, and there is no system-level benefit to using them. However, we have seen that plasmonic modulators can be very energy-efficient running at tens of GHz. These modulators have an intrinsically-low Q that allows for temperature tolerance. Of course a low-Q typically is associated

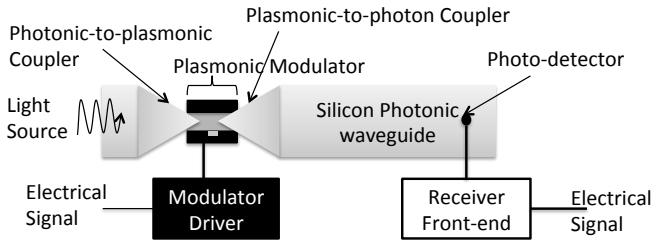


Fig. 3. Hybrid Link: By using a plasmonic modulator and silicon photonic waveguide, we can achieve the best of both worlds: long range propagation, and low power consumption and high performance modulation. Couplers convert photons into SPP and vice versa.

with poor signaling and overall low channel bandwidth due to limited number of wavelengths that can be supported on the waveguide. In this section, we will discuss the design of our novel hybrid photonic/plasmonic interconnect architecture, starting with a new hybrid photonic/plasmonic link, extending it to operate under wavelength-division multiplexing, and examining the trade-offs in variation tolerance it presents.

A. Single-Wavelength Hybrid Photonic/Plasmonic Channel

We propose to use energy-efficient high-speed plasmonic modulators in conjunction with conventional photonic waveguides. This hybrid channel will benefit from the best of both technologies, i.e. low-loss waveguiding of photonics and high-speed energy-efficient modulation of plasmonics. Moreover, using a plasmonic modulator with intrinsically low-Q will improve the temperature and process variation tolerance of the whole channel. Figure 3 shows the details of the proposed hybrid photonic/plasmonic channel. An off-chip laser source generates a continuous-wave (CW) laser that is exciting SPP in a very short (around 4 μm) plasmonic waveguide using a photonic-to-plasmonic coupler, that couples photons into SPP. These SPP are modulated using a plasmonic modulator which is derived by an electrical circuit. SPPs are coupled back into photons using a plasmonic-to-photonic coupler. These photons propagate in the photonic waveguide and finally are detected using a photo detector that needs a TIA to convert the photo-current into a voltage.

It is worth noting that this design, as described, does not support WDM because of the fact that the modulator is in the way of the signal and although WDM in plasmonic modes is theoretically possible, it has not been demonstrated yet. In the next subsection, we discuss how to add WDM support to this hybrid channel without losing temperature tolerance properties. In order to evaluate the energy efficiency of this hybrid channel, we compare the energy cost of a single wavelength hybrid channel to that of photonic and electrical channels.

a) Electrical Channels Power Estimation: We use McPat 0.8 [27] to estimate the power consumption of electrical RC links, assuming 22 nm low standby power transistors. We limited the frequency to 3 GHz to keep the router power consumption in check. Higher frequencies consume much more than 1 W per router, a prohibitive amount of power for a scalable many-core interconnect. For this link-level analysis we assume an activity factor of 100% for each of the components, although the effect of this factor is explored in a more detail during our detailed link level design space exploration in Section V.

Plasmonic Modulator Insertion Loss	3 dB
Photonic-to-Plasmonic Coupler	1.1 dB
Photo Detector	1 dB
Photonic Modulator insertion loss	0.1 dB
Photonic Ring Drop Loss	0.16 dB
Photonic Ring Through Loss	0.001
Waveguide loss	0.1 dB/mm
Waveguide Bending Loss	0.005
Fiber-to-waveguide loss	0.5 dB
Photonic waveguide splitter loss	0.04
Laser efficiency	30%

TABLE II
PHOTONIC LOSSES USED IN THE MODEL TO CALCULATE LASER STATIC POWER CONSUMPTION.

b) Photonic Channels Power Estimation: A photonic link consists of an off-chip laser source, transmitter, waveguide and a receiver as shown in Figure 1. Power consumption is divided between static laser power, static thermal-tuning power, and dynamic power consumption depending on the number and distribution of transmitted bits. Laser power is estimated by adding all optical losses along the path and using the detector sensitivity. Thermal tuning is assumed to consume around 50 μW/°C [11] and assuming a temperature variance of 20°C, we estimate a 1 mW per ring thermal tuning power. This is in accordance with the 100 fJ/bit heating power assumed in [3].

Table II shows the losses used in our energy calculations. Transmitters are ring resonator modulators derived by an electric circuit and receivers are a Ge photo-detector connected to a TIA and an amplifier. Using parameters of electrical components at 22 nm technology from [3], we estimate that electrical components will consume 88 fJ/bit. The modulator is estimated to consume 50 fJ/bit at 10 Gbps as in [38].

c) Hybrid Channel Power Estimation: We model all optical losses including the plasmonic-to-photonic coupler losses, modulator insertion loss and waveguide losses. We use the same electrical components as [3] except for the modulator driver, we estimate that it will consume 10 fJ/bit since it is driving a low-capacitance modulator. We assume the compact modulator [7] and a 1.1 dB coupler loss per transition [10]. While these assumptions are somewhat optimistic, a sensitivity analysis of their implications is discussed in Section VI. Although plasmonic modulators can run at many tens of Gbps, we assume an electrically limited system with 10 Gbps bandwidth.

Figure 4 shows the energy per bit consumption against the length of the link for all three configurations: electrical RC wire, photonic point-to-point channel and the proposed hybrid photonic/plasmonic channels. Hybrid channels reduce the distance at which photonics becomes more energy efficient than electrical from 300 μm to around 200 μm. This energy saving results from two sources: energy-efficient modulation and the reduction of thermal tuning requirement under certain process variation assumptions as discussed in Section IV-C. Even if the energy-efficiency of photonic modulators can be improved to be around 5 fJ/b, this should not change our

conclusion that the hybrid channels are more efficient than photonic ones, due to more tolerance of variation.

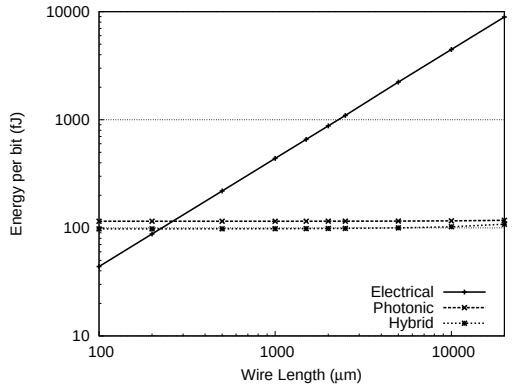


Fig. 4. Energy per bit against link length for different technologies: electrical, photonic and hybrid. It is clear that electrical signaling is more efficient for any link of less than 500 μm length. Beyond that, hybrid links are the most energy efficient.

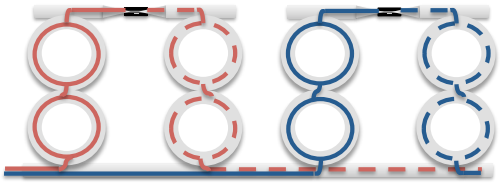


Fig. 5. Hybrid channel with wavelength-division-multiplexing support: Passive ring filters are used to divert a certain wavelength from the main waveguide to an auxiliary hybrid channel in which modulation is done.

B. Hybrid Channel with WDM support

While we believe we are the first to propose the use of this form of hybrid plasmonic/photonic link structure, it has a significant drawback – it works for only a single wavelength. Given the size of the photonic waveguides, we will be far better off if we can compress multiple logic channels onto a single waveguide. The channel architecture we propose to support WDM, couples a micro-ring resonator as a passive filter to remove a certain wavelength from the main waveguide and drop it on an auxiliary waveguide. This auxiliary waveguide can be the hybrid channel discussed in Section IV-A. Figure 5 shows an example of how such design can support WDM. The red wavelength is filtered out of the channel and then coupled into plasmonic modes using a plasmonic-to-modulator coupler and then modulated using a plasmonic modulator and then coupled back into a photonic mode. Finally, it is filtered from the auxiliary waveguide back to the main waveguide. This can be repeated for many wavelengths and thus, WDM is supported.

The main advantage of using the hybrid channel is the reduction of the thermal tuning requirement of photonic modulators, but we are now introducing rings back into the design. How can we keep the improved variation and temperature tolerance provided by single-wavelength hybrid channels but still operate in WDM? Somewhat counterintuitively, we propose a system can work with a special *low-Q ring filter* to decrease the coupling ratio between the ring and the waveguide. Typically, photonics engineers struggle to achieve high-Q, as a low-Q ring usually means both lower modulator performance and lower number of channels (wavelengths) which are limited by the free-spectral-range (FSR). However, with a plasmonic modulator these two aspects are now decoupled. The FSR can be increased by reducing the radius of the ring, but as we reduce the number of the bending loss become more dominant and the insertion loss increases. A trade-off between the coupling ratio and ring radius will determine the drop-loss of the filter, 3-dB bandwidth and 20-dB bandwidth. *In essence, this new design allows us to make a new trade-off between bandwidth and temperature/ process variation sensitivity.*

We did an analysis of these effects and the results are shown in Figures 6-8. The x-axis in all of these figures is the coupling ratio between the waveguide and the ring filter (higher coupling ratio means high-Q ring). Figure 6 presents the 3-dB bandwidth (or Full-Width at Half Maximum (FWHM) bandwidth) for rings of radius from 1 to 2μm. It shows that the higher the coupling ratio the lower the bandwidth of the ring, thus the more temperature dependent the filter is. Figure 7 shows the drop loss of the filter as a function of coupling ratio. It is worth noting that higher coupling ratio and higher-Q rings leads to lower insertion loss and lower radius rings increases the insertion loss due to higher bending losses. Figure 8 provides the 20-dB bandwidth that can be used along with channel spacing to determine the number of channels that this filter supports based on -20 dB crosstalk. Channel spacing should be at least half of 20-dB bandwidth to support -20 dB crosstalk. Thus, the number of channels increases as we increase the coupling ratio. This gives us a trade-off between overall channel bandwidth and temperature and variation tolerance.

C. Temperature and variation tolerance discussion

The hybrid channel process and temperature variation tolerance depend on variation tolerance of both plasmonic modulators and micro-ring filters. Since plasmonics is a very young field, there is not, to the best of our knowledge, any experimental study on how process variation affects plasmonic modulators. However, having an intrinsic low-Q encourages us to think it would tolerate more errors. We will limit the following analysis to the passive ring filter used in the hybrid channel.

Temperature and process variation tolerance depend on three parameters: 3-dB bandwidth, 20-dB bandwidth and channel spacing. Large 3-dB bandwidth allows for channel deviation without missing the signal, however, this deviation is limited by the crosstalk to a neighbor channel (which is determined by channel spacing and 20-dB bandwidth). We can define variation tolerance in one direction by:

$$V = \min(0.5 \times \text{3-dB BW}, \text{channel spacing} - 0.5 \times \text{20-dB BW})$$

In order to study the variation tolerance of passive filters, we setup an experiment after [17] in which we model two process variation effects, a systematic one that affects all rings and a local effect that vary from one ring to another. We capture these effects by generating a Gaussian random number with standard deviation from a range of values determined by experimental studies. Tuning is done according to a slightly

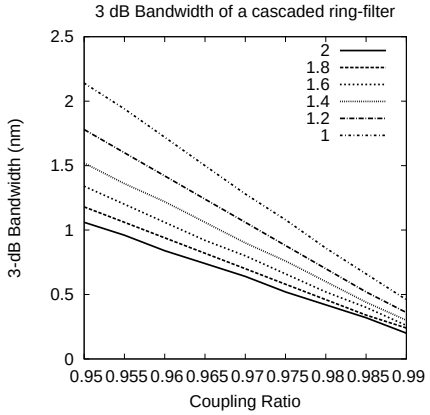


Fig. 6. 3-dB Bandwidth of ring filter against coupling ratio for different ring radii.

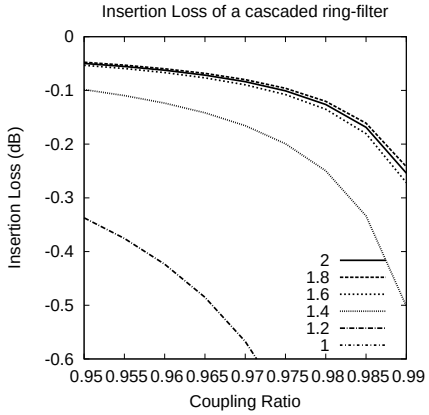


Fig. 7. Insertion loss as a function of coupling ratio for different ring radii.

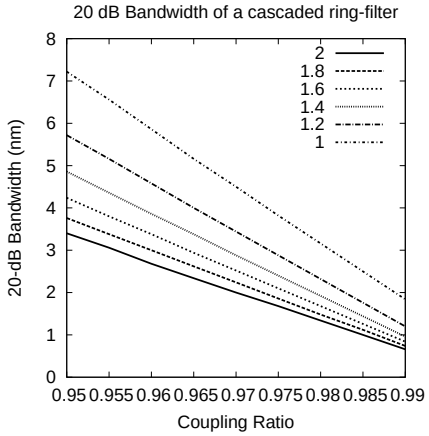


Fig. 8. 20-dB Bandwidth as a function of coupling ratio for different ring radii.

modified scheme from the one devised by Georgas rt al. [17] in which heating is only used to tune rings to a neighbor channel (by one or more hops in one direction towards the red) and the original ordering of the bits is restored using a barrel shifter (we ignored the barrel shifter power consumption in all our experiments). We assume a 44μW/°C heater efficiency and 0.09 nm/°C change in resonance due to temperature [33]. We did an experiment with systematic variation standard deviation of 0.08 nm and a local variation standard deviation between 0 nm and 7 nm. For all coupling ratios, radii and channel configurations, we generated 1000 trials and calculated the power consumption used for tuning the rings. It is worth noting that there are other tuning techniques such as current injection that will lead to a reduction in the power consumption but its tuning range is limited and we believe that both low and high-Q rings will benefit from using it. Due to space limitation, we chose to show results for only rings of 1.4 μm radius used to build 4 and 32 channels due to its moderate insertion loss (0.16 dB) for low-Q filters.

Figure 9 and 10 show a color map of the tuning power consumption for a channel built using 4 and 32 channels, respectively, as we vary the local variation standard deviation from 0 to 7 nm for different coupling ratios. For 4-channels, it is clear that a low-Q ring (97% coupling ratio) consumes much less tuning power (lighter color) than a high-Q ring (99% coupling ratio) for all variation ranges. This is not the case for 32-channels in which low-Q filters have a slightly higher advantage at low variation levels (below 1 nm) vs. high variation levels. This is because the distance that a ring has to be moved to a neighbor channel is reduced by the smaller channel spacing of the 32-channels as compared to the 4-channels which reduces the benefits gained of using a low-Q ring. This shows a new trade-off between process variation tuning power and area efficiency.

HP Labs measurements indicate that resonance wavelength shifts by $\pm$0.4 nm across a die and by $\pm$2 nm across wafer [16]. This is also confirmed by other studies showing a standard deviation of 0.4 [49], 0.68 nm [34], 0.75 nm [28] and 1 nm [46]. This means that low-Q rings will provide an advantage over high-Q micro-ring filters.

It is worth noting that our optimization is possible because we use rings as passive filters. Micro-ring modulators need to have Q in the range of 10,000 to be able to switching quickly in the range of 10-20 Gbps [45]. Hence, we cannot have a low-Q photonic ring modulator which makes our approach interesting because the hybrid channel can still support WDM (although with lower number of wavelengths) while running the plasmonic modulator at high speed. PMMA-cladding [47] has been proposed to athermalize the photonic micro-ring resonators; however, this athermalization reduces the fact that heaters can be used to correct resonance-shift due to process variations. Another proposal is to use Mach-Zehneder interferometer adjacent to the ring filter to provide a temperature-tolerance range of 80°C [19]. However, such technique negates any area efficiency of the micro ring resonators.

V. Link Level Design Space exploration

We just presented a filter-level analysis of process variation tolerance of the low-Q rings. A link-level design space exploration is required to verify our hypothesis on a system-level and quantify the conditions under which the hybrid channel can save power. The hybrid channel provides two new trade-offs: bandwidth density vs. power consumption and laser power consumption (based on passive filter drop loss) vs. tuning power consumption (depending on coupling ratio).

In order to study these trade-offs, we performed an analysis based on the tuning power consumption calculated for different filters in Section IV-C. We assumed a channel of 15 mm length that can handle 640 Gbps traffic and 100% activity factor. We used different configurations described in Table III to implement such channel. Hybrid and photonic channels use the parameters used for the experiment in Section IV-A except for the ring insertion loss which is derived from that data presented in Figure 7. Hybrid channels use two filters for each modulator, and one filter for a detector per wavelength. Photonic channels use one ring modulator and one filter for a detector per wavelength. Modulators are assumed to be of 10 Gbps bandwidth and channels are built using one or more waveguides to achieve the required bandwidth based on the limitation of wavelengths per waveguide of each configuration. We used data from the 2 μm radius with 99% coupling ratio filter (Q around 7750) to model the photonic modulators.

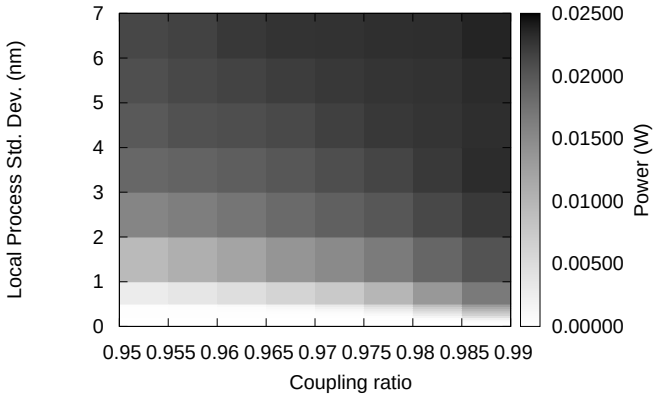


Fig. 9. Tuning power consumption for 4-channels using micro-rings of 1.4 μm radius as we vary local variation (standard deviation) and different coupling ratio (systematic variation std. dev. is 0.08 nm). Higher coupling ratio means higher-Q ring.

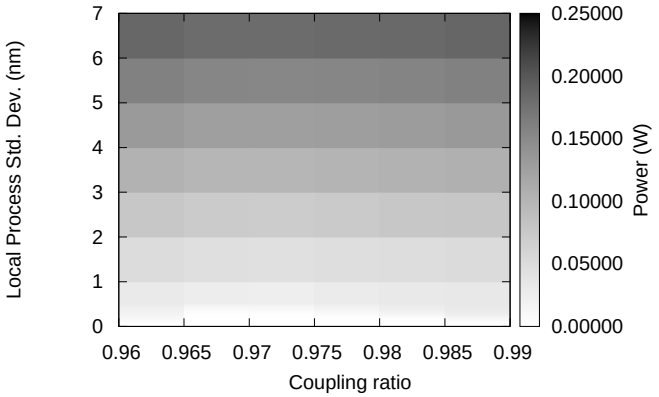


Fig. 10. Tuning power consumption for 32-channels using micro-rings of 1.4 μm radius as we vary local variation (standard deviation) and different coupling ratio (systematic variation std. dev. is 0.08 nm). Higher coupling ratio means higher-Q ring.

Channel Type	Filter radius (μm)	Coupling ratio	Wavelengths per waveguide	Area relative to photonic with 64 wavelengths
Hybrid	1.0	95%	4	16x
Hybrid	1.0	95%	16	4x
Hybrid	1.0	96%	32	2x
Hybrid	1.4	95%	4	16x
Hybrid	1.4	95%	16	4x
Hybrid	1.4	96.5%	32	2x
Photonic	2.0	99%	16	4x
Photonic	2.0	99%	32	2x
Photonic	2.0	99%	64	1x

TABLE III
THE DIFFERENT CONFIGURATION OF FILTERS USED IN TESTING THE ENERGY COST UNDER DIFFERENT PROCESS VARIATION SCENARIOS.

Figure 11 shows energy cost per transmitted bit normalized to the energy consumption of photonic channel with 64-wavelengths of each cluster (its energy cost is labeled in fJ/bit over its bar). Energy cost is divided into three categories (laser, dynamic and tuning). Results are shown for these different configurations at various levels of local process variation standard deviation of 0.0, 0.5, 1.0 and 2.0 nm and systematic local variation standard deviation of 0.08 nm. Bar labels are formatted channel type(ring radius, wavelengths per waveguide). For low-levels of local variation (0.0 standard deviation), there is very little energy saving due to tuning. However, for moderate levels of variation around 0.5 nm local variation standard deviation, we can see that although laser power consumption of the hybrid channel utilizing a filter of radius 1.4 μm is higher than that of the photonic channels using the high-Q modulators, the overall energy cost of the hybrid channel can save up to 45% of the most bandwidth density efficient photonic channel of 64 wavelengths per waveguide. This of course comes at the cost of bandwidth density, as this channel will occupy 4 times the area of the photonic one. We believe that this new trade-off will give system designers a new knob to use in their optimization. For higher levels of variation (local variation standard deviation of 1.0), the hybrid channel with 1 μm radius filter and 4 wavelengths per waveguide, consuming 8 times laser power consumption saves 28% of the most area efficient configuration. This shows that the design space is huge and needs to be carefully examined according to the experimental studies of process variation. It is worth noting that laser power consumption is an off-chip one while tuning is done on-chip. Finally, hybrid channels cannot save energy for high levels of process variation (as in the case of local variation std. dev. of 2 nm) because even if the filter requires slightly less tuning power as shown in last section, the fact that we need 3 filters (rather than just 2 in the photonic channel) outweighs any filter based benefits. This is the same reason hybrid channels with 32-wavelengths consume more energy than its photonic counterparts under 0.5 and 1 nm local variation standard deviation. However, initial experimental measurements indicate that local process variations are within 0.5-1 nm [16], [46], [34], [49], [28].

Both laser and tuning energy consumption are static that are not dependent on activity, however, we want to check whether the activity factor can affect the relative advantage of any of these schemes or not. We plotted energy cost against activity factor for all configurations assuming local variation standard deviation of 0.5 nm in Figure 12. The relative advantage of a configuration over another is maintained regardless of the activity except for one case between one photonic channel with 16 wavelengths and one hybrid channel with 32 wavelengths. The photonic channel is more efficient for activity factors higher than 0.85. This is not the case in the hybrid channels because the static power consumption of laser and tuning is dominant.

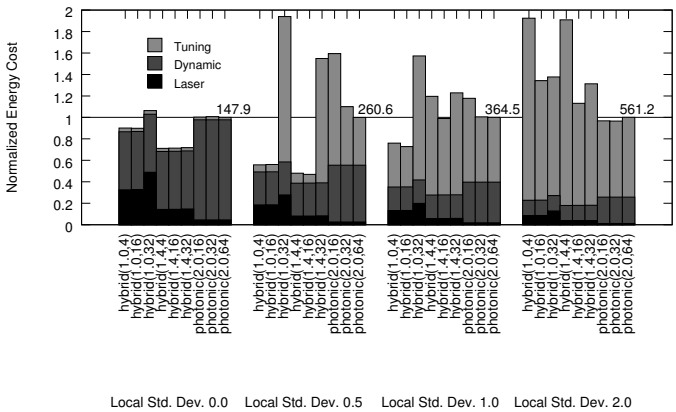


Fig. 11. Energy cost of different channel configurations carrying 640 Gbps under a systematic process variation standard deviation of 0.08 nm and different local process variation. Each bar is normalized to the photonic channel with 64 wavelengths in each cluster which is indicated in fJ/bit above its bar.

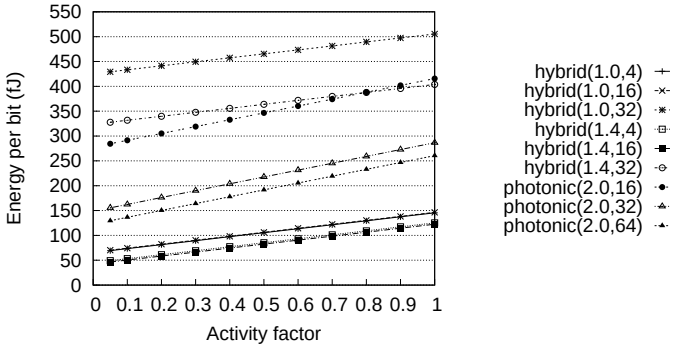


Fig. 12. Energy cost as a function of activity factor for different channel configuration under a systematic process variation standard deviation of 0.08 nm and local process variation standard deviation of 0.5 nm (electrical wires energy cost as a function of activity factor was calculated but it was never the best option for this wire length and was not shown to preserve clarity of the other configurations. For example, at 5% activity factor it consumes 7x the energy of the most efficient optical configuration, and at 50% activity the ratio goes up to more than 40x.).

VI. Architectural Implications

Hybrid channels differ from ring-modulator based photonic channels in two main things: the insertion loss of the plasmonic modulator can be very high and the limited number of wavelengths on the waveguide. Plasmonic modulator insertion loss was estimated to be from 1 dB [13] to 12 dB. This high range of variation affects the static laser power consumption depending on the topology in which the channel is used. In this section we characterize the architectural implications of using the hybrid photonic/plasmonic channel for optical buses, configurable meshes and point-to-point channels.

A. Optical Buses

There are two main configurations for optical buses that can be used to build optical crossbars. The first one is Multiple-Writer-Single-Reader (MWSR) in which one node listens to certain wavelengths and all other nodes can send to this node. An example of this topology is Corona [43]. In a ring-based implementation, only the writer tunes its modulator to on-resonance mode and all other nodes' modulators are tuned off-resonance so that the signal passes by on the waveguide. Unfortunately, we cannot utilize the same technique in our channel because our ring filters are passive, hence the signal will be filtered out of the main waveguide even if the node is not sending and it will pass through the plasmonic modulator (on on-state) and back to the main waveguide. This means that the signal will suffer from n-1 modulator insertion losses, 2n-1 ring drop losses, 2w(n-1) ring through losses and 2n plasmon-to-photonic coupler losses, where n is the number of nodes on the bus and w is the number of wavelengths on channel. Figure 13 presents a contour plot of the laser static power as we vary the modulator insertion loss and the number of nodes on the crossbar (assuming 32 wavelengths can be sent on the waveguide, 12 cm waveguide and 7 bends) in order to evaluate the range of values for which the hybrid channel can be used to build a MWSR bus. Optical losses are modeled using parameters in Table II. The x-axis shows the number of nodes on the bus and the y-axis represents the different values for a plasmonic modulator loss. It shows that even for very optimistic modulator insertion loss, the maximum number of nodes to be connected with 20 W static power budget is 7 nodes. Although that seems to be a small number for network-on-chip applications, it can still be used for memory buses.

The second type of optical bus-based crossbars is Single-Writer-Multiple-Reader(SWMR) bus was first proposed by Kırman et al. [24]. In a conventional implementation, the sender broadcasts a message on another control bus declaring the destination of its message so that all other nodes to detune their nodes. Since we are restricting our channel to passive ring filters, we will have to use splitters to split the signal along the way to multiple destinations. We assume a 0.4 dB loss 5-95% splitter for the 95% branch to be used at each receiver. Each node will suffer from one modulator insertion loss, one ring drop losses, 2w ring through loss and 95% loss (13 dB), where n is the number of nodes on the bus and w is the number of wavelengths on channel. We assumed the same length and bends of the MWSR analysis.

Figure 14 shows contour lines of laser power as we change the number of nodes in the crossbar and insertion loss of the plasmonic modulator. The axes are the same as in Figure 13. For a laser power budget of 20 W, 20 nodes are possible for 1dB insertion loss and 10 nodes can be built having a modulator with IL of 4 dB. This shows that the SWMR is more tolerant to modulator IL than MWSR. However, it still does not scale to networks of tens of nodes that are required for any network-on-chip in a high performance chip.

B. Meshes and Tori

There have been two main proposals in the architecture community of optical meshes and tori. Electrically-controlled circuit switched optical torus net was proposed as a high bandwidth optical interconnect [40]. Having a circuit switched network makes it possible to use active low-Q filters (requires more switching time) because the circuit can amortize the increased time of setting the circuit up. However, it is highly unlikely that a circuit switched network can be used for cache-line transfers that tend to be small and short in life time.

Phastlane was proposed as an optical mesh in which a control packet is sent a cycle ahead of the data to configure the router rings in order to divert the signal to the correct port [8] which requires very aggressive switching times of active ring filters; so that the path is ready when data signal arrives a few ps later. Even if we allowed active switching of the ring filters used in the hybrid channel, they cannot operate fast because of their low-Q because it takes more time and energy to shift the resonance range outside the 1 nm bandwidth. Hence, we conclude that it will be extremely unlikely to build such configurable meshes using the hybrid channel design.

C. Point-to-Point Channels

Many architectural proposals suggested the use of photonics in point-to-point channels [26] and [20]. In a point-to-point channel, the signal will suffer from 1 modulator insertion loss, 2(w - 1) through-loss and 1 drop loss, where w is the number of wavelengths in the channel. Figure 15 studies the effect

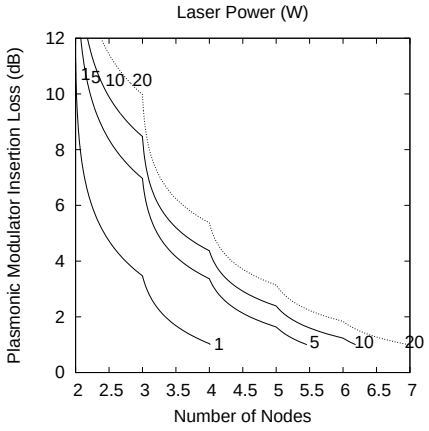


Fig. 13. MWSR: Sensitivity analysis of laser power (W) required for a bus of n nodes and 32 wavelengths as a function of plasmonic modulator insertion loss.

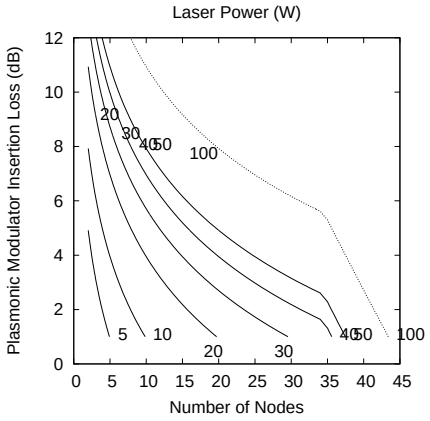


Fig. 14. SWMR: Sensitivity analysis of laser power (W) required for a bus of n nodes and 32 wavelengths as a function of plasmonic modulator insertion loss.

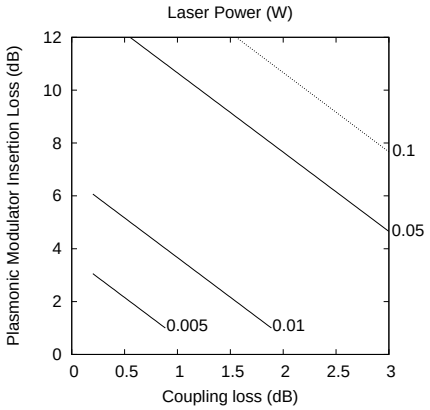


Fig. 15. Point-to-point: Sensitivity analysis of laser power required for point-to-point channel as a function of modulator insertion loss and photonic-to-plasmonic coupler loss.

of plasmon-to-photon coupling loss and modulator insertion losses on laser power of a two end-points channel assuming 32 wavelengths channels. Even under very pessimistic conditions of 10 dB insertion loss and 2.5 dB coupling loss, the channel power will be 0.1 W. To put things in perspective, a 64-node (8,8,8) Clos network has a total of 128 inter-router channels, which means under these pessimistic assumption 12.8 W laser power which is still acceptable given that laser is an off-chip resource. Under more favorable assumptions, total laser power can be 1.28 W which is extremely low. The fact that laser power is not high in point-to-point channels even under pessimistic conditions stems from the low number of photonic components and simplicity of the channel design.

Considering that point-to-point channels are the most suitable for our hybrid channel, we will evaluate it from a system point of view against optical and electrical networks in the next section.

VII. Evaluation

In order to understand system level implications of the hybrid channel, we chose to evaluate high-radix low-diameter networks. We use these links in the context of a flattened butterfly and clos topologies which reduce the average number of hops through the use of high radix routers [21], [39]. Specifically, we evaluate the following configurations:

1) 64-node electrical 2D flattened butterfly (flatfly-electrical);
2) 64-node hybrid 2D flattened butterfly (flatfly-hybrid);
3) 64-node photonic clos network (8,8,8) (clos-photonic); and
4) 64-node hybrid clos network (8,8,8) (clos-hybrid)

A flattened butterfly topology is derived from a butterfly topology in which each row of routers is collapsed into one high-radix router keeping inter-stage links [22]. This results in a direct topology with high bisection bandwidth. When mapped into an on-chip network, it forms a matrix of routers in which each row, as well as each column, of routers are fully connected. This results in channels that span one full dimension of the processor (approximately 15 mm in a 400 mm^2 chip). It is an opportunity for the distance-independent power consumption provided by hybrid plasmonic/photonic links. A clos network is a 3-stage network in which each input router is connected to all middle routers and a middle router is connected to all output routers. We use the layout proposed by [20] in which each row of routers is clustered in a router group resulting in 8 router groups that are connected to each other using 128 channels.

In order to make a fair comparison between hybrid channels and photonic channels, we assumed a hybrid channel that has half of the bandwidth of a photonic channel taking into consideration that a photonic channel area is half that of a hybrid one because hybrid waveguides can only carry 32 wavelengths instead of 64. It is worth noting that for long channels end-points (consisting of rings and auxiliary waveguides) account for only just 2% of the channel area given the 5 μm pitch and a channel of 5 cm long. If area is not a concern, e.g. in a case of 3D integration, both should be able to achieve same bandwidth. We assumed that electrical signaling is a 128-bit bus running at 3 GHz.

A. Experimental Setup

We used a heavily modified version of Booksim 2.0 [9] to provide performance results using synthetic traffic patterns. The simulator is warmed up until steady state is reached and statistics are reset, then a sample of the packets is measured from the time it enters the source queue until it is received. Simulation runs until all packets under measurements leave the network. Table IV shows the simulation parameters.

We also model the losses indicated in Table II. The simulator power calculations were done using McPat 0.8 [27]. We assumed a 22 nm process with low standby power transistors running at 3 GHz. We compute power consumption by running synthetic traffic patterns through the simulator and gather activity factors of channels and routers to calculate dynamic power consumption. In the flattened butterfly, we evaluate a 64 core network in which each set of four cores are connected to a router. Having 16 routers organized in a 4x4 matrix, electrical channels between routers are assumed to be 2, 4 and 6 cycles (spanning 5, 10 and 15 mm). It is worth noting that electrical channels are assumed to be pipelined every 2.5 mm but that any flip-flop power consumption is neglected.

Parameter	Photonic	Electrical and Hybrid
Flit size	256 bit	128 bit
Flits per packet	1	2
Virtual channels	1	1
Buffers Per VC	4	4
Router radix	8x8(clos) and 10x10(FBfly)	
Router delay	4 cycles	4 cycle
SW and VC allocator	iSLIP [30]	iSLIP [30]
Routing algorithm	rand. middle (Clos) and random_min (FBfly)	
CPU Frequency	3 GHz	3
Transistor type	LSPT	LSPT

TABLE IV
SIMULATION PARAMETERS USED IN OUR NOC STUDY.

Moreover, core-to-router channels are neglected for all four configurations. In a flattened butterfly, hybrid point-to-point channels (carrying 32 wavelengths) are assumed to be of 1 cycle latency (because of 10.45 ps/mm propagation delay and a cycle time of more than 300 ps). The longest link of 15 mm has a propagation delay of 156.75 ps which will leave around 150 ps for modulation and detection latency. Having a matrix layout mandates crossings between the waveguides, however a recent multi-layer waveguide technology suggests that we get can get away from crossings [5] so we do not include them in our loss model.

In a clos layout, we assume that the maximum waveguide length is 5 cm (using two laser sources, we assume conservatively 2 cycles for all channels) and we assume that a waveguide can carry 64 wavelengths. Splitter losses are ignored for all hybrid and photonic configurations.

B. Latency

While the latency of the individual links is quite low, when considered in the context of a full network, the buffer and routing delays become very important. In order to see the effect of hybrid links on latency, We present results for only uniform and bitcomp traffic patterns (due to page limitations) for injection rates from 0.001 to saturation in Figures 16 and 17. Comparing the two flattened butterfly configurations, it is clear that hybrid channels provide lower zero-load latency and supports higher saturation bandwidth than its electrical counter-part because of its ability to serve more packets. Clearly, photonic-clos saturates at double the injection rate of the hybrid-clos because of the increased bandwidth in photonic links but they have similar trends in zero-load-latency. In many multi-core systems, network latency and power consumption are the primary design requirements, even more so than throughput [39].

As can be seen, the hybrid links provide consistently lower end-to-end latency over electrical, with improvements around 40% under uniform traffic pattern, mostly due to the faster long links. While these latency reductions are helpful, a breakdown of the power shines a light on the potential of these hybrid channels. Clos has higher saturation bandwidth because it has higher bisection bandwidth than flattened butterfly topology.

C. Power

In our final set of experiments we try to evaluate the trade-off between the reduction of tuning power and static laser power of the hybrid and photonic channels as well as the power savings against electrical channels. We used injection rates of 5% to calculate power consumption because all configurations were below saturation at this injection rate. Figure 18 shows power consumption break down for the uniform random, transpose, shuffle, bitcomp and bitrev traffic patterns for the photonic and hybrid clos configurations under different variation assumptions. We have categorized the power into 5 bins: buffer power, router power (including the switch itself), static electrical power, dynamic channel power (the power actually consumed pushing bits across the chip), laser power (generated off chip but accounted for in the total budget including generation inefficiency), and heating power required for rings in photonic and hybrid channels. For the first two bars in each cluster, we assumed that there is no process variation and 20°C temperature variation range. Photonic ring thermal tuning is optimistically assumed to be 5μW/°C for 20°C (the state of the art is 50μW/°C [11]) and hybrid channels can be assumed to be without any tuning power consumption (if PMMA-cladding is used thermal sensitivity is just 0.027nm/°C leading to more than 40°C tolerance). In that case, hybrid channels save on average 27% of total network power consumption depending on the traffic pattern. The next four bars in each cluster look at two different ranges of variation (either due to process variation or temperature): a standard deviation of variation of 0.5 nm and 6 nm, respectively. We calculate the tuning power using the experiments presented in Section V using 1.4μm rings with 97% coupling ratio and 2.0 μm rings with 99% coupling ratio for low and high-Q rings, respectively. We used the tuning power of 32-channels for the low-Q filter and 64-channel for the high-Q modulator. Although the tuning power consumption is lower for hybrid channels in both cases, the savings effect is much higher in total network power consumption in the case of the high variation because it dominates the power consumption. It is worth noting that this result is for different channel bandwidth requirements and two channels with similar bandwidth would use approximately the same tuning power for such large deviation as in the 6 nm case depending on the chosen ring radius and channel density as discussed in Section V.

Figure 19 compares the electrical and hybrid channel implementations of the Flattened butterfly. Hybrid channels save on average 93% of the channel power consumption and 64% of the total network power against the electrical signaling under no process variation. This comes from the fact that the power consumption of hybrid channels is almost completely distance independent while the electrical ones are not. Even under variation, the savings are on average 61% and 34% for variation standard deviation of 0.5 nm and 6.0 nm, respectively.

VIII. CONCLUSIONS AND FUTURE WORK

There is little doubt that photonic interconnects will push deeper and deeper into the network hierarchy, and already today we are beginning to see commercial applications of "optical USB" and other board-level technologies. Bringing these photonic technologies down to the scale that they can be integrated with a single chip is a challenging problem, but one that has received serious attention as of late. There are several major challenges that must be overcome, including temperature dependence and manufacturing variability. Current known

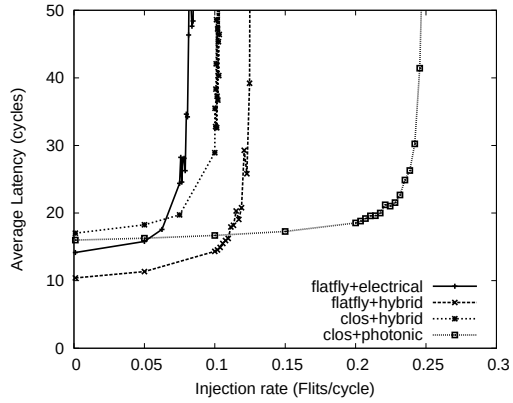


Fig. 16. Average latency against injection rate for uniform traffic pattern.

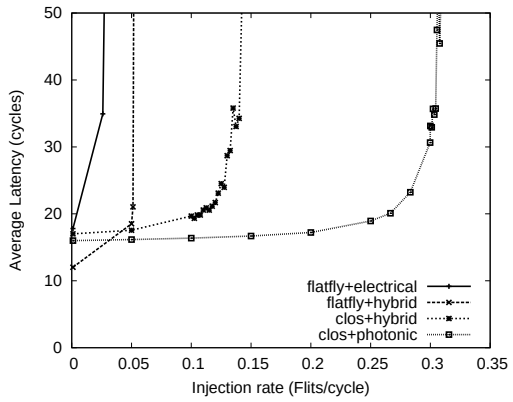


Fig. 17. Average latency against injection rate for bitcomp traffic pattern.

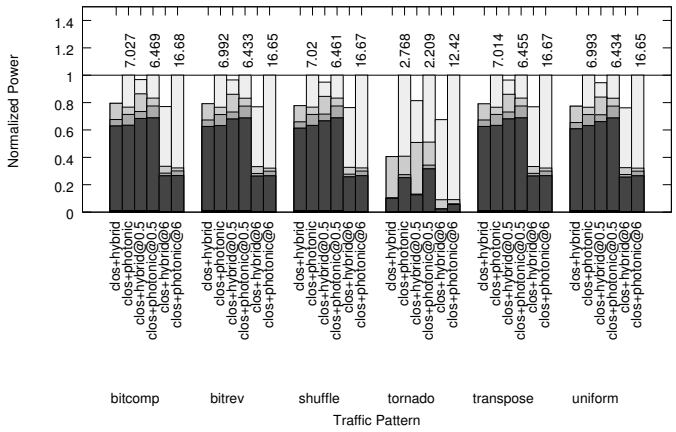


Fig. 18. Normalized power consumption to the photonic configuration for different variation conditions @ 5% injection rate for hybrid and photonic Clos. The number after the "@" indicates the variation std. dev. in nm.

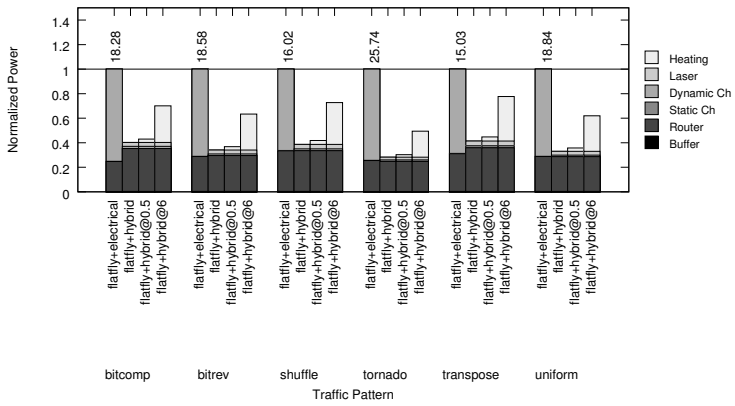


Fig. 19. Normalized power consumption to the electrical configuration @ 5% injection rate for electrical and hybrid Flattened Butterfly. The number after the "@" indicates the variation std. dev. in nm.

methods of dealing with these issues require heating elements to keep the rings precisely calibrated within 0.15 nm for a ring modulator with Q of 10,000 [41]. This means that a change more than 2°C will put this modulator out of resonance. Unfortunately, photonic ring modulator has to have a Q of 10K or 20K to be able to switch fast. Moreover, athermalization techniques solve the temperature tolerance problem for 20°C variance, but it cannot solve the process variation resonance shift problem.

Plasmonic devices offer an interesting new approach to this problem, primarily because plasmonic modulators offer high speed operation with low-Q, making them much more tolerant of temperature variation. While our analysis suggests that pure plasmonic interconnects with MDM waveguides cannot currently compete directly with electrical networks due to high loss (signals are dampened below the point of recovery after only a few tens of micrometers), when used in conjunction with optical waveguides they present an interesting new opportunity. The inherently temperature independent behavior of the plasmonic modulators, coupled with the low-Q of passive rings, allow us to improve static power consumption significantly by designing low-Q passive ring filters to be used with plasmonic modulators which can reduce total energy cost per bit from 28% to 45% of the most area efficient photonic channel under low to moderate variation. Our preliminary results in this study should motivate plasmonics researchers to study the thermal and variation dependence of plasmonic modulators to verify the intuition behind this paper experimentally. It is worth noting that the laser power is an off-chip resource which has little effect on the temperature of the die, while thermal tuning is done on-chip.

Of course, as with all technologies, there is a trade-off. Our new hybrid photonic/plasmonic channel has a significantly higher insertion loss making it incompatible with scalable full-crossbar designs. However, for point-to-point channels it can tolerate up to 12 dB modulator insertion loss with 0.05 W laser which is still lower 0.064 W for thermal tuning assuming the same number of rings. Furthermore, our new approach makes the trade-off between bandwidth (number of wavelengths on the channel) and temperature sensitivity explicit.

This is an exciting time for the architecture community to engage the plasmonics community, as they are actively pushing the devices in many new directions, like using plasmonics to implement a photo-transistor that will eliminate the need for the TIA. Moreover, plasmonic nano-laser may lead to another way of direct modulation on-chip. Many papers have discussed new ways by which SPPs can be propagated over longer and longer distances, but to our knowledge this is the first paper to seriously consider trade-offs in their use in on-chip links, how they can be coupled with photonic structures to make new network primitives, and the classes of interconnect topologies

that are feasible under their use.

ACKNOWLEDGMENTS

The authors would like to thank Martijn Heck and the anonymous reviewers for providing useful feedback on this paper.

REFERENCES

[1] H. A. Atwater. The promise of plasmonics. *Scientific American*, 296(4):56–63, April 2007.

[2] J. Balfour and W. J. Dally. Design tradeoffs for tiled cmp on-chip networks. In *ICS '06: Proceedings of the 20th annual international conference on Supercomputing*, pages 187–198, 2006. ACM.

[3] C. Batten, A. Joshi, J. Orcutt, A. Khilo, B. Moss, C. W. Holzwarth, M. A. Popovic, H. Li, H. I. Smith, J. L. Hoyt, F. X. Kartner, R. J. Ram, V. Stojanovic, and K. Asanovic. Building many-core processor-to-dram networks with monolithic cmos silicon photonics. *IEEE Micro*, 29(4):8–21, 2009.

[4] S. Beamer, C. Sun, Y.-J. Kwon, A. Joshi, C. Batten, V. Stojanović, and K. Asanović. Re-architecting dram memory systems with monolithically integrated silicon photonics. *SIGARCH Comput. Archit. News*, 38(3):129–140, 2010.

[5] A. Biberman, N. Sherwood-Droz, X. Zhu, M. Lipson, and K. Bergman. High-speed data transmission in multi-layer deposited silicon photonics for advanced photonic networks-on-chip. CLEO'11, 2011.

[6] M. Brongersma. personal communication, 2010.

[7] W. Cai, J. White, and M. Brongersma. Compact, high-speed and power-efficient electrooptic plasmonic modulators. *Nano Letters*, 9(12):4403–4411, 2009.

[8] M. J. Cianchetti, J. C. Kerekes, and D. H. Albonesi. Phastlane: a rapid transit optical routing network. In *ISCA '09: Proceedings of the 36th annual international symposium on Computer architecture*, pages 441–450, 2009. ACM.

[9] W. Dally and B. Towles. *Principles and Practices of Interconnection Networks*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

[10] C. Delacour, S. Blaize, P. Grosse, J. M. Fedeli, A. Bruyant, R. Salas-Montiel, G. Lerondel, and A. Chelnokov. Efficient directional coupling between silicon and copper plasmonic nanoslot waveguides: toward metaloxidesilicon nanophotonics. *Nano Letters*, 10(8):2922–2926, 2010.

[11] C. T. DeRose, M. R. Watts, D. C. Trotter, D. L. Luck, G. N. Nielson, and R. W. Young. Silicon microring modulator with integrated heater and temperature sensor for thermal control. In *Lasers and Electro-Optics (CLEO) and Quantum Electronics and Laser Science Conference (QELS), 2010 Conference on*, pages 1 –2, may 2010.

[12] J. Dionne, L. Sweatlock, M. Sheldon, A. Alivisatos, and H. Atwater. Silicon-based plasmonics for on-chip photonics. *Selected Topics in Quantum Electronics, IEEE Journal of*, 16(1):295 –306, jan.-feb. 2010.

[13] J. A. Dionne, K. Diest, L. A. Sweatlock, and H. A. Atwater. Plasmostor: A metal-oxide-si field effect plasmonic modulator. *Nano Letters*, 9(2):897–902, 01 2009.

[14] J. A. Dionne, L. A. Sweatlock, H. A. Atwater, and A. Polman. Plasmon slot waveguides: Towards chip-scale propagation with subwavelength-scale localization. *Physical Review B (Condensed Matter and Materials Physics)*, 73(3):035407+, 2006.

[15] P. Dong, S. Liao, D. Feng, H. Liang, D. Zheng, R. Shafiiha, C.-C. Kung, W. Qian, G. Li, X. Zheng, A. V. Krishnamoorthy, and M. Asghari. Low vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator. *Opt. Express*, 17(25):22484–22490, Dec 2009.

[16] M. Fiorentino. Talk at HSD 2011, 2011.

[17] M. Georgas, J. Leu, B. Moss, C. Sun, and V. Stojanovic. Addressing link-level design tradeoffs for integrated photonic interconnects. In *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, pages 1 –8, sept. 2011.

[18] J. Grandidier, G. C. des Francs, S. Massenot, A. Bouhelier, L. Markey, J.-C. Weeber, C. Finot, and A. Dereux. Gain-assisted propagation in a plasmonic waveguide at telecom wavelength. *Nano Letters*, 9(8):2935–2939, August 2009.

[19] B. Guha, B. B. C. Kyotoku, and M. Lipson. Cmos-compatible athermal silicon microring resonators. *Opt. Express*, 18(4):3487–3493, Feb 2010.

[20] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic. Silicon-photonic clos networks for global on-chip communication. In *Proceedings of the 2009 3rd ACM/IEEE International Symposium on Networks-on-Chip*, NOCS '09, 2009.

[21] J. Kim, J. Balfour, and W. Dally. Flattened butterfly topology for on-chip networks. In *MICRO 40: Proceedings of the 40th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 172–182, 2007.

[22] J. Kim, W. J. Dally, and D. Abts. Flattened butterfly: a cost-efficient topology for high-radix networks. In *ISCA '07: Proceedings of the 34th annual international symposium on Computer architecture*, pages 126–137, New York, NY, USA, 2007. ACM.

[23] J. Kim, W. J. Dally, S. Scott, and D. Abts. Technology-driven, highly-scalable dragonfly topology. In *Proceedings of the 35th Annual International Symposium on Computer Architecture*, ISCA '08, 2008.

[24] N. Kirman, M. Kirman, R. K. Dokania, J. F. Martínez, A. B. Apsel, M. A. Watkins, and D. H. Albonesi. Leveraging optical technology in future bus-based chip multiprocessors. In *MICRO 39: Proceedings of the 39th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 492–503, 2006.

[25] N. Kirman and J. F. Martínez. A power-efficient all-optical on-chip interconnect using wavelength-based oblivious routing. In *ASPLOS '10: Proceedings of the fifteenth edition of ASPLOS on Architectural support for programming languages and operating systems*, 2010.

[26] P. Koka, M. O. McCracken, H. Schwetman, X. Zheng, R. Ho, and A. V. Krishnamoorthy. Silicon-photonic network architectures for scalable, power-efficient multi-chip systems. In *Proceedings of the 37th annual international symposium on Computer architecture*, ISCA '10, 2010.

[27] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi. Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO 42, 2009.

[28] Z. Li, M. Mohamed, X. Chen, E. Dudley, K. Meng, L. Shang, A. Mickelson, R. Joseph, M. Vachharajani, B. Schwartz, and Y. Sun. Reliability modeling and management of nanophotonic on-chip networks. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 20(1):98 –111, jan. 2012.

[29] K. MacDonald and N. Zheludev. Active plasmonics: current status. *Laser & Photonics Reviews*, 4(4):562–567, 2010.

[30] N. McKeown. The islip scheduling algorithm for input-queued switches. *IEEE/ACM Trans. Netw.*, 7(2):188–201, 1999.

[31] F. J. Mesa-Martinez, E. K. Ardestani, and J. Renau. Characterizing processor thermal behavior. In *Proceedings of the fifteenth edition of ASPLOS on Architectural support for programming languages and operating systems*, ASPLOS '10, 2010.

[32] T. Nikolajsen, K. Leosson, and S. I. Bozhevolnyi. Surface plasmon polariton based modulators and switches operating at telecom wavelengths. *Applied Physics Letters*, 85(24):5833–5835, 2004.

[33] C. Nitta, M. Farrens, and V. Akella. Addressing system-level trimming issues in on-chip nanophotonic networks. In *High Performance Computer Architecture (HPCA), 2011 IEEE 17th International Symposium on*, feb. 2011.

[34] J. S. Orcutt, A. Khilo, C. W. Holzwarth, M. A. Popović, H. Li, J. Sun, T. Bonifield, R. Hollingsworth, F. X. Kärtner, H. I. Smith, V. Stojanović, and R. J. Ram. Nanophotonic integration in state-of-the-art cmos foundries. *Opt. Express*, 19(3):2335–2346, Jan 2011.

[35] D. Pacifici, H. J. Lezec, and H. A. Atwater. All-optical modulation by plasmonic excitation of cdse quantum dots. *Nature Photonics*, 1(7):402–406, July 2007.

[36] Y. Pan, J. Kim, and G. Memik. Flexishare: Energy-efficient nanophotonic crossbar architecture through channel sharing. In *IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2010.

[37] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary. Firefly: illuminating future network-on-chip with nanophotonics. In *ISCA '09: Proceedings of the 36th annual international symposium on Computer architecture*, 2009.

[38] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. J. Thomson. Silicon optical modulators. *Nature Photonics*, 4(8):518–526, July 2010.

[39] D. Sanchez, G. Michelogiannakis, and C. Kozyrakis. An analysis of on-chip interconnection networks for large-scale chip multiprocessors. *ACM Trans. Archit. Code Optim.*, 7(1):1–28, 2010.

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

[41] N. Sherwood-Droz, K. Preston, J. S. Levy, and M. Lipson. Device guidelines for wdm interconnects using silicon microring resonators. In *Workshop on the Interaction between Nanophotonic Devices and Systems (WINDS 2010)*, dec. 2010.

[42] D. Vantrease, N. Binkert, R. Schreiber, and M. H. Lipasti. Light speed arbitration and flow control for nanophotonic interconnects. In *MICRO 42: Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, 2009.

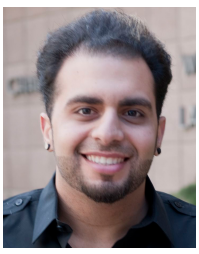
[43] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn. Corona: System implications of emerging nanophotonic technology. In

ISCA '08: Proceedings of the 35th Annual International Symposium on Computer Architecture, 2008.
[44] G. Veronis and S. Fan. Theoretical investigation of compact couplers between dielectric slabwaveguides and two-dimensional metal-dielectric-metal plasmonicwaveguides. *Opt. Express*, 15(3):1211–1221, 2007.
[45] Q. Xu, D. Fattal, and R. G. Beausoleil. Silicon microring resonators with 1.5-μm radius. *Opt. Express*, 16(6):4309–4315, Mar 2008.
[46] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson. 12.5 gbit/s carrier-injection-based silicon micro-ring silicon modulators. *Opt. Express*, 15(2):430–436, Jan 2007.
[47] 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*, 21(17):1175 –1177, sept.1, 2009.
[48] R. Zia, M. D. Selker, P. B. Catrysse, and M. L. Brongersma. Geometries and materials for subwavelength surface plasmon modes. *J. Opt. Soc. Am. A*, 21(12):2442–2446, 2004.
[49] W. A. Zortman, D. C. Trotter, and M. R. Watts. Silicon photonics manufacturing. *Opt. Express*, 18(23):23598–23607, Nov 2010.

Hassan M. G. Wassel is pursuing a Ph.D. in the Department of Computer Science at UC Santa Barbara. His research focuses on computer architecture and its interaction with software and novel hardware technologies in order to build high performance, secure, and energy efficient systems. He has a B.Sc. and M.Sc. from the Department of Computer Science and Automatic Control at Alexandria University (Egypt) in 2004 and 2007, respectively. He is a student member of the IEEE and the ACM.

Daoxin Dai (M07) received the B. Eng. degree from the Department of Optical Engineering, Zhejiang University, Hangzhou, China, and the Ph.D. degree from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2000 and 2005, respectively. Then, he joined Zhejiang University as an assistant professor and became an associate professor in 2007, a full professor in 2011. His current research interests include silicon micro-/nano-photonic integrated devices. He has authored and coauthored about 90 refereed international journals papers. He is a member of IEEE.

Mohit Tiwari is a Computing Innovation Fellow at University of California, Berkeley. He received his PhD in Computer Science from University of California, Santa Barbara in 2011. His research uses computer architecture and program analyses to build secure, reliable systems, and has received a Best Paper award at PACT 2009, an IEEE Micro Top Pick in 2010, and the Outstanding Dissertation award in Computer Science at UCSB in 2011.



Jonathan K. Valamehr is pursuing a Ph.D. in the department of Electrical and Computer Engineering at the University of California, Santa Barbara in Santa Barbara, California. His research focuses on using hardware and emerging technologies in computer architecture to develop secure microprocessors. He received B.S. (2003) and M.S. (2007) degrees in Electrical and Computer Engineering, both from the University of California, Santa Barbara. He is a student member of the IEEE and the ACM.

Luke Theogarajan (M02) received the Bachelor degree in electronics and telecommunications engineering from the Manipal Institute of Technology, India, in 1994, the M.S. degree from Arizona State University in 1996 and the Ph.D. degree from Massachusetts Institute of Technology (MIT), Cambridge, in 2007. He is currently an Assistant Professor of electrical and computer engineering at the University of California, Santa Barbara. His research interests include combining the processing power of electronics with the versatility of synthetic chemistry to develop neural prosthetic devices and using optical interconnects for energy efficient computing. He is a recipient of the NIH New Innovator award (2010) the NSF CAREER award. He holds four patents.

Jennifer Dionne received the M.S.(2005) and Ph.D. (2009) degrees in applied physics from Caltech, and the B.S. (2003) degree in physics and systems science and engineering from Washington University in St. Louis. She is currently an Assistant Professor of Materials Science and Engineering at Stanford University. As a graduate student, her research focused on plasmonic structures for subwavelength waveguides, active electro-optic devices, and visible frequency negative index materials. Her current research interests include electrooptical properties of colloidal nano-crystals for applications ranging from renewable energy to biosensing. Dr. Dionne was named one of Technology Reviews TR35 (2011), and is a recipient of the NSFCAREER award (2012).

Frederic T. Chong is the Director of Computer Engineering and a Professor of Computer Science at UCSB. He also directs the Greenscale effort in Energy-Efficient Computing, which involves over 20 multi-disciplinary faculty. Chong received his Ph.D. from MIT in 1996 and was a faculty member and Chancellor's fellow at UC Davis from 1997-2005. He is a recipient of the NSF CAREER award and his research interests include emerging technologies for computing, multicore and embedded architectures, computer security, and sustainable computing.

Timothy Sherwood is an Associate Professor of Computer Science specializing in high performance and high assurance embedded systems. On four separate occasions his work has been selected by IEEE Micro as a "Top-Pick", he is a recipient of an NSF Career Award, and winner of the Northrup Grumman Excellence in Teaching Award. Prior to joining UCSB, he graduated with a B.S. in Computer Science from UC Davis (1998), and M.S. and Ph.D. from UC San Diego (2003).