This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

A Survey of Silicon Photonics for Energy Efficient Manycore Computing

Sudeep Pasricha (sudeep@colostate.edu), Mahdi Nikdast (mahdi.nikdast@colostate.edu)
Colorado State University, Fort Collins, CO

Abstract: Silicon photonics has emerged as an exciting new paradigm to reduce the overheads of data movement in manycore computing platforms. However, many open problems remain to be addressed before energy-efficient and high-performance optical communication can be realized at the chip-scale. This article surveys the landscape of solutions across the device, circuit, and architecture layers, as well as cross-layer techniques to enable energy-efficient data movement over optical interconnects for on-chip and off-chip communication subsystems in manycore computing platforms.

Keywords: silicon photonics, networks-on-chip, on-chip communication, chip-to-chip communication, energy efficiency, process variations, thermal variations, reliability

1 Introduction

By the early 2000s, the limits of instruction-level parallelism (ILP) had become apparent to chip architects who were attempting to improve processor performance to meet the growing demands of increasingly complex workloads. New deeper and wider pipelined processor architectures designed around that time to maximize ILP came at a cost: high power dissipation, which state-of-the-art chip packaging and cooling solutions could not handle. This "power wall" forced chip designers to abandon power hungry ILP designs in favor of thread-level parallelism (TLP) and data-level parallelism (DLP), which relied on multiple simpler processor cores on a chip with single-instruction-multiple data (SIMD) or vector instruction support to increase parallelism and workload performance. The resulting change in processor design brought forth new challenges: designers quickly realized that communication among multiple cores, and between cores and memory was now a performance and energy bottleneck. To put it simply, moving data to and from a processor now took more time and energy than computing with the data in the processor. This observation ushered in the era of communication-centric chip design, with innovations such as networks-on-chip (NoCs) receiving significant research interest to reduce the data movement overheads. The growth of the IEEE Hot Interconnects Symposium (HOTI) [1] and the IEEE/ACM Network-On-Chip Symposium (NOCS) [2] over the past two decades are indicative of the many challenges and open problems with data movement in computing, which are even today being addressed by a vibrant community of researchers and practitioners.

State-of-the-art mainstream computing systems today have manycore general-purpose processor chips with tens of cores (*e.g.*, AMD EPYC processor family with up to 64 cores [3] and Intel's Xeon Platinum processor family with up to 56 cores [4]) connected using a NoC architecture and up to 8 sockets of these chips interconnected together. Emerging GPUs and neuromorphic accelerator chips with hundreds to thousands of cores are now further pushing the boundaries of on-chip and off-chip communication architecture design. For example, NVIDIA's GPU chips with the Turing architecture have more than 4000 CUDA cores [5] and AMD's Navi/RDNA GPU architecture supports more than 2500 cores [6]. As another example, Cerebras recently unveiled an artificial intelligence (AI) accelerator processor chip with 1.2 trillion transistors and 400,000 (lightweight) cores [7]. While such a chip may not be representative of commercially-viable mainstream processors, it points to a future where hundreds to thousands of CPU, GPU, and AI cores will need to be connected together with high bandwidth and low power interconnect solutions.

Indeed, energy for data movement is one of the biggest challenges in computing today. On a modern processor chip fabricated in CMOS technology, it takes 0.1-0.2 pJ/bit for transfers over a 1 mm long electrical link (*e.g.*, for a core to access an L2 cache bank), and 1-4 pJ/bit over longer electrical links (*e.g.*, for a core to access larger and more distant last level L3 caches). When going off-chip to access main memory (DRAM), it can take up to 30 pJ/bit. Inter-socket (chip-to-chip) transfers can take 11-16 pJ/bit with AMD's Infinity Fabric [8] and Intel's UltraPath Interconnect Fabric [9]. While these numbers may seem small, they are at least 100× larger than the projected communication energy budget for computing substrates in the future, to meet the goals for the US Department of Energy (DOE) Exascale supercomputing initiative that aims to achieve exaflop performance within a 20 MW overall power budget [10]. Even energy for processing data reveals interesting insights about the role of data movement: it takes about 1.7 pJ/bit to perform a floating-point operation in a modern processor [11], but a standard-cell-based, double precision fused-multiply add (DFMA) requires only ≈20 pJ, which reveals that fetching operands is much more energy-consuming than computing on them.

The fundamental issue with data movement today lies with the electrical interconnect technology in use for on-chip and off-chip communication, where limits of electron movement in conductors constrain the energy and latency of transfers [12]. Taking inspiration from the datacom and telecom domains where long-haul (several kilometers long) optical communication has been in use since the late 1970s [13], researchers began to explore the possibility of using optical interconnects in computing platforms. The seminal work by Goodman *et al.* [14] advocated for the use of optical interconnects in VLSI chips as far back as 1984 to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

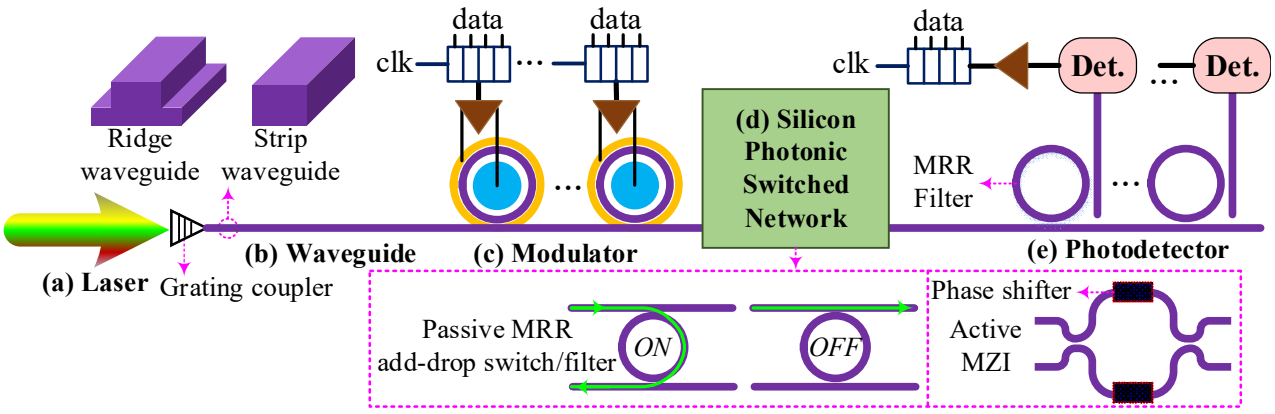


Figure 1: An abstract overview of a silicon photonic link, including (a) a laser source, which can be off-chip or on-chip; (b) waveguide, which can be a strip waveguide (for passive devices and networks) or a ridge waveguide (for active devices and networks); (c) modulator, which modulates the electronic data ('data' in figure) on an optical signal (here we consider a microring resonator modulator as an example); (d) photonic switched network, which includes many basic switching elements (*e.g.*, those based on microring resonators (MRRs) or Mach-Zehnder interferometers (MZIs)) responsible for switching and routing optical signals from a source towards a destination; and, (e) photodetector, which detects the optical signal and converts it to the original modulated data and is often connected to a filter (MRR filter in figure) to drop the wavelength of interest to the corresponding photodetector.

reduce latency. But it was really in the early to mid-2000s, with the advent of the communication-centric chip design era and the rise of silicon photonics that interest in optical interconnects at the chip-scale truly picked up steam. Advances in silicon photonics have enabled integrated photonic circuits with the fabrication of photonics devices that use silicon as the optical medium, and employ existing semiconductor fabrication techniques [15]. As silicon is already used as the substrate for most CMOS integrated circuits (ICs), researchers realized that it was possible to create chips in which the optical and electronic components could be integrated together. A new vision of computing emerged where data processing is done with electrons while data movement is achieved with photons. Over the past 15 years, this vision has slowly but surely been on the path to realization. Today rack-to-rack and board-to-board optical interconnect solutions have already been commercialized and are in use [16]-[18]. Chip-to-chip and on-chip optical interconnects are also actively being explored in industry and academia. Such optical communication within and between multicore chips will be essential to overcoming the data movement challenge in emerging and future computing platforms. However, chip-scale optical interconnects still face several daunting challenges that increase their energy footprint, which weakens the case for using chip-scale optical interconnects to replace electrical ones.

In this article, we survey the landscape of design innovations and architectures to enable energy-efficient manycore computing with silicon photonics. We begin by reviewing the state-of-the-art with silicon photonics devices. We then focus our attention on approaches to improve energy efficiency at the device, circuit, and architecture levels that can enable the design of energy-efficient optical interconnects and networks at the chip-scale. We then discuss the role of cross-layer techniques that may span across one or more of the device, circuit, architecture, and system layers to achieve energy-efficient communication with optics technology. We conclude with a discussion on open challenges that require further research and industry focus.

2 Overview of Silicon Photonic Devices

Over the past decade, different CMOS-compatible silicon photonic devices have been developed to realize chip-scale communication in manycore computing platforms [152]. Fig. 1 shows an abstract overview of some fundamental silicon photonic devices required in photonic interconnects, including a laser source, waveguides, modulators, switching elements, and photodetectors. We briefly review the state-of-the-art and energy constraints of each of these devices in the rest of this section.

Lasers: The light source (see Fig. 1a) can be off-chip or integrated on a chip. Off-chip lasers have high light-emitting efficiency and good temperature stability at a cost of large coupling losses (and hence large energy consumption) between the off-chip light source and the silicon chip, which is mostly due to the grating coupler loss, as well as higher packaging costs. On the other hand, on-chip lasers can potentially achieve a higher integration density and better performance in terms of energy efficiency [19]. Nevertheless, the development of on-chip lasers on silicon is extremely challenging because of the low emission efficiency of silicon [20], and its sensitivity to chip-wide thermal variations. Yet, several on-chip silicon lasers have been proposed: III-V-based silicon lasers via bonding techniques [21], quantum dot (QD) lasers monolithically grown on silicon [22], and germanium-on-silicon lasers for large-scale monolithic integration [70]. Using a hybrid III-V (Indium Phosphide) active layer, [142] developed a distributed feedback (DFB) laser on a SiO_2/Si substrate and reported a direct modulation rate of 25.8 Gbps and an energy efficiency of 0.5 pj/bit at a chip stage temperature of 25-50 °C. A hybrid QD laser with non-return-to-zero (NRZ) communication at a record high direct modulation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

rate of 15 Gb/s with energy efficiency of 1.2 pJ/bit was proposed in [22], which operates at chip stage temperatures of up to 70°C. The performance of each of these on-chip laser contenders should be assessed in terms of operating wavelength, pump condition, power consumption, fabrication process and cost, and thermal stability, which is a crucial parameter to consider in complex optoelectronic integrated circuits and photonic interconnects.

Waveguides: Silicon photonic waveguides are the electrical wire counterparts in integrated silicon photonic circuits, as shown in Fig. 1b. In general, one can classify optical waveguides as strip waveguides and ridge waveguides. Strip waveguides are usually used in passive devices and circuits, in which light is passively routed (*e.g.*, in wavelength-switched networks). Ridge waveguide are often employed in active devices and circuits, in which light is routed actively through, for example, electro-optic or thermo-optic effects. Ridge waveguides allow for electrical connections to be made to the waveguides for active tuning of silicon photonic devices (*e.g.*, through P-N junctions in a modulator). The high refractive index contrast in silicon-on-insulator (SOI) waveguides, in which the silicon core is often buried in silica (*i.e.*, SiO_2), allows for sub-micron waveguide dimensions (*e.g.*, 500×220 nm) with high efficiency [23]. Indeed, light propagation in photonic interconnects relies significantly on precise geometry adjustment of optical waveguides. Therefore, any distortion in waveguide geometries and shape can notably impact the optical power and energy efficiency in waveguides. For instance, sidewall roughness due to inevitable lithography and etching process imperfections can result in scattering, and hence optical losses in waveguides. Such optical loss is often defined as dB/cm (*i.e.*, propagation loss) and typically ranges from 0.5 dB/cm to 2 dB/cm in single-mode SOI waveguides [24]. In addition to propagation loss, there is power loss whenever a waveguide bends (*i.e.*, bending loss). This bending loss, which is proportional to the bending curvature, is a result of radiation and mode-mismatch in waveguide bends, and can be as around 0.01 dB for a 5 μm 90° bend [33]. Compared to the high-refractive-index-contrast SOI platform, silicon nitride can provide an alternative CMOS-compatible waveguide solution with a moderate index-contrast [25]. In particular, the propagation loss in silicon nitride waveguides can be potentially an order of magnitude lower: Silicon nitride waveguides with losses smaller than 1 dB/m have been proposed, but this comes at the expense of lateral and/or vertical confinement [26]-[28].

Modulators: The gateway from electronic data to an optically modulated signal is through silicon photonic modulators, as indicated in Fig. 1c. In general, modulators fall into two categories: those based on direct absorption, which usually only operate over a limited wavelength range close to the absorption edge of the material and provide amplitude modulation but not independent phase modulation, and those relying on embedded phase shifters capable of supporting both complex-valued modulation and optical broadband operation [29]. Since the first demonstration of a silicon photonic modulator with gigahertz modulation frequencies in 2004 [30], there has been a noticeable amount of research effort on improving modulation efficiency, bandwidth, and insertion losses, all affecting energy efficiency in photonic interconnects. In particular, electro–optic (E/O) cutoff frequencies in excess of 50 GHz have been reached with the carrier depletion modulators [31], and even better performance with speeds in excess of 100 GHz has been presented [32]. Mach-Zehnder Interferometers (MZIs) [34] and microring resonators (MRR) [35] are the two common silicon photonic devices widely used for designing modulators, each with certain benefits and energy costs. When designing modulators, trade-offs have to be made when aiming at fast devices that feature simultaneously low drive voltage and small footprint. For example, carrier-injection (active) devices feature voltage-length products as small as $V_\pi L = 0.36$ Vmm [36], where V_π is the voltage required to achieve a phase shift equals to π and L is the length of the device, but at a cost of limiting the modulation speed to a few Gb/s. In contrast, carrier-depletion modulators support higher modulation rates, but typical voltage-length products are beyond 10 Vmm [36]. Although drive voltage and device footprint can be substantially reduced by using MRR structures [35], these devices feature limited optical bandwidth and their frequency response is prone to temperature fluctuations, which will be discussed in the next section. In [36], a modulator with semiconductor hybrid (SOH) polymer integration suitable for operation at 40 Gb/s was demonstrated with a phase shifting efficiency smaller than 0.5 Vmm and energy efficiency of 0.4 pj/bit. Recently, a silicon MRR with integrated thermo-optic resonance tuning was proposed in [143] with modulation data rate up to 128 Gb/s (64 Gbaud PAM4), thermo-optic phase efficiency of 19.5 mW/π-phase shift, and energy efficiency of 18 fj/bit.

Switches: Silicon photonic switching elements are the primary building blocks in chip-scale switched photonic networks-on-chip [37]. In general, we can divide switching elements into two categories (see Fig. 1d): passive switching elements, in which an optical signal is passively routed based on its optical wavelength (*e.g.*, wavelength-selective routing [38]), and active switching elements, in which an optical signal is actively routed through thermo-optic or electro-optic effects [39] in silicon. Moreover, the switching mechanism can be in terms of resonant effects (*e.g.*, in MRRs) or based on interference effects (*e.g.*, in MZIs). Accordingly, compact silicon photonic switching elements based on MZIs and MRRs have been developed [39]. Power loss, crosstalk, bandwidth, switching speed, and extinction ratio are a few important parameters in silicon photonic switching elements. Actively switching of light can be performed through either thermo-optic effects (*e.g.*, thermal tuning) or electro-optic effects (*e.g.*, current injection). Compared with electro-optic switching elements, thermo-optic switching elements often present better power efficiency at a cost of lower switching speed (few microseconds versus few nanoseconds in electro-optic switches). Recently, a silicon photonic switching element has been developed based on micro-electromechanical systems (MEMS)-actuated adiabatic couplers with an extinction ratio as low as -70 dB, worst-case insertion loss better

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

Table 1. An overview of the state-of-the-art optical devices discussed in Section 2. Such devices are required in an on-chip communication link in a photonic network-on-chip.

Device	Technology	Performance	Bandwidth / Data rate	Reference
On-chip laser	Hybrid III-V (Indium Phosphide) on silicon	Energy efficiency: 0.5 – 0.6 pj/bit	25.8 Gbps	[142]
Grating coupler (for off-chip laser)	Silicon on Insulator (SOI)	Coupling efficiency: 81% (-0.9 dB)	38.8 nm	[146]
Waveguide	220 nm SOI	Loss: <0.25 dB/cm	-	[147]
	220 nm silicon nitride	Loss: <0.1 dB/cm	-	[147]
Modulator	MRR with integrated thermo-optic resonance tuning (PAM4)	Energy efficiency: 18 fj/bit	128 Gb/s	[143]
Switch	1x2 MEMS-actuated adiabatic coupler (SOI platform)	Through loss: 0.02 dB Drop Loss: 0.47 dB Extinction ratio: 70 dB ON voltage: 30 V OFF voltage: 20 V ON switching time: 0.72 μs OFF switching time: 0.78 μs	300 nm	[40]
Optical filter	MEMS tunable MRR-based add-drop filter	Static power: <100 nW Resonance wavelength tuning: 530 pm	20 Ghz	[146]
Photodetector (PD)	Pin waveguide PD with hetero-structured silicon-Ge-silicon (Si-Ge-Si) junction	Leakage current: 150 nA Bias voltage: -0.5 V Sensitivity: 11.25 dBm	25 Gbps	[42]

than 0.5 dB, sub-microsecond switching speed, and a high bandwidth of up to 300 nm [40].

Optical filters: The continuous growth in bandwidth requirements in photonic NoCs has driven a trend of integrating multiple optical wavelengths (channels) into these networks. Indeed, different optical wavelengths, each carrying a separate modulated optical signal, can simultaneously propagate through a single waveguide without interaction (*i.e.*, wavelength division multiplexing). As a result, there is a need for an optical filter that can efficiently filter out and distinguish among different optical wavelengths at the receiver (see Fig. 1e). Silicon photonic MRRs are good candidates for optical filtering because of their wavelength selectivity [144]. In an MRR-based add-drop filter, an MRR in proximity with two waveguides (see Fig. 1d and 1e) can filter (drop) an optical wavelength that matches with the MRR resonance wavelength from one waveguide to the other. The power loss and bandwidth of the MRR determines the optical filtering efficiency [144]. Multiple MRRs can be coupled together to create a high-order MRR filter with better out-of-band rejection ratio and pass band [145]. Recently, [146] reported an add-drop filter using a microelectromechanical (MEMS) tunable MRR with 530 pm resonance wavelength turning and static power consumption below 100 nW, but with a low bandwidth of 20 Ghz [145]. A comprehensive analysis and review of MRR-based optical filters is provided in [144].

Photodetectors: The transparency of silicon in wavelength bands near 1310 nm and 1550 nm makes it an excellent choice of material for low-loss chip-scale communication. However, for the very same reason, silicon is inefficient for detection. Hence, silicon photonic photodetectors (shown in Fig. 1e) require integration of other materials (*e.g.*, III-V or germanium) for high speed and efficient photonic detection [41]. Silicon photonic photodetectors require large bandwidth, high efficiency, and low dark current. There are often numerous design trade-offs between speed, efficiency, and output power. For example, designing for high bandwidth favors small devices, but at a cost of lower output power [41]. Indeed, the photodetector output power (responsivity: electrical output per optical input) is an important parameter to be considered when designing photonic interconnects. It determines the minimum optical signal power level required at the photodetector to correctly detect the modulated data on the signal (*e.g.*, logic one or zero). Therefore, the sum of optical losses on a photonic link should be always smaller than the photodetector responsivity. Recently, [42] reported a high performance pin waveguide photodetector made of a lateral hetero-structured silicon-Ge-silicon (Si-Ge-Si) junction operating under low reverse bias at 1550 nm. This photodetector shows efficiency-bandwidth products of ≈9 GHz at −1 V and ≈30 GHz at −3 V, with a leakage dark current as low as ≈150 nA. Moreover, it achieves a bit-error rate of 10^{-9} for conventional 10 Gbps, 20 Gbps, and 25 Gbps data rates, yielding optical power sensitivities of −13.85 dBm, −12.70 dBm, and −11.25 dBm, respectively [42].

Summarizing the fundamental devices discussed in this Section, Table 1 reviews the state-of-the-art of various optical devices required in an on-chip communication link in a photonic NoC and lists critical performance parameters for each device.

3 Device Challenges and Enhancements

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

This section reviews some of the fundamental challenges at the silicon photonic device-level in terms of vulnerability to different variations and aging, which impact the energy efficiency of silicon photonic devices, and some state-of-the-art solutions to overcome such challenges.

Process Variations

There have been many successful demonstrations of the silicon photonic devices discussed in the previous section, to pave the way for their integration into manycore computing systems. All of such devices, however, are susceptible to inevitable variations in fabrication process. As mentioned before, any variation in the critical dimensions (*e.g.*, waveguide thickness or width) of a silicon photonic waveguide can considerably impact the device performance [43]-[46], often in terms of imposing higher losses and crosstalk noise, both of which impact the energy efficiency of silicon photonic devices. Such device inefficiencies can accumulate in a system, considerably degrading the system performance [47]. For example, it was shown that fabrication process variations can decrease the optical signal-to-noise ratio (OSNR) in photonic interconnects by up to 20 dB [47]-[48]. Recent efforts have proposed design solutions to improve the tolerance of silicon photonic devices to fabrications process variations [49]-[53]. For example, [51] proposed to improve the physical design parameters (*e.g.*, waveguide width) in MRRs and improved MRR tolerance to fabrication process variations by more than 60%.

Thermal Variations

In addition to fabrication variations, silicon photonic devices are sensitive to runtime temperature fluctuations. In particular, this is of concern when such devices are integrated and packaged with electronic devices, where chip-scale temperature variations can reach up to 30 degrees [54]. Indeed, silicon refractive index is temperature dependent (due to the thermo-optic effect) and follows $n = n_0 + \frac{dn}{dt}\Delta T$, where $n_0$is the refractive index at room temperature, dn/dT is the thermo-optic coefficient of silicon that is in a range of $1.8\text{x}10^{-4}\text{K}^{-1}$ [55], and ΔT is the chip temperature variation. As a result, similar to the impact of fabrication variations, temperature variations can significantly degrade the performance of photonic interconnects [56]-[59] because of the high thermo-optic effect in silicon. For example, [56] indicated a considerable increase in system power loss in photonic interconnects due to runtime temperature variations. There have been some efforts at the device level to improve thermal stability of silicon photonic devices, such as passive temperature stabilization using liquid crystals [60] and development of athermal solutions (*e.g.*, based on polymers and titanium dioxide) for silicon photonic devices [61], [62]. For example, through the use of organically modified sol-gel claddings, [61] demonstrated a thermal shift down to −6.8 pm/°C for transverse electric (TE) polarization in MRRs with waveguide widths of 325 nm.

Impact of Variations

Due to the fabrication and thermal variations discussed above, devices such as MRRs are prone to resonance wavelength drifts, which prevents accurate modulation, switching, and filtering for detection. For example, [45] indicated that with a single 1 nm change in the waveguide thickness due to fabrication variations, the resonant wavelength of an MRR can be shifted by at least 2 nm. Such deviations impose energy efficiency degradation in photonic interconnects especially when using multiple wavelengths in photonic interconnects. The number of wavelengths used per waveguide is often referred to as the degree of wavelength division multiplexing (WDM), with each wavelength enabling the transfer of a stream of bits, in parallel with other wavelengths, to support high bandwidth transfers. The proposed solutions to address variation-induced shifts fall into two main categories: permanent post-fabrication trimming and runtime tuning mainly through the thermo-optic effect (*i.e.*, thermal tuning) or electro-optic effect (i.e., bias or current injection tuning).

Overcoming Impact of Variations

The main post-fabrication trimming solutions are based on either changing the level of compaction or stress of the cladding or core material (*e.g.*, using high-energy electron or laser beams), or changing the refractive index of the cladding material by applying high-energy UV light. Thermal tuning is achieved by varying current through a heater near the MRR, causing an increase in the refractive index of the silicon and the resonant wavelength to red shift. The current injection tuning method injects (or depletes) free carriers into (or from) the Si core of an MRR using an electrical tuning circuit, which reduces (or increases) the MRR's refractive index owing to the electro-optic effect, to compensate for the variation-induced red (or blue) shift in the MRR's resonance wavelength. Current injection tuning can provide a tuning range of only 1.5 nm at most [63], but it incurs relatively low latency and power overheads (an addition of up to 130 μW/nm to the total link power [64]). In contrast, thermal tuning incurs high latency and power overheads (an addition of 550 mW/nm shift, to the total link power [65], and in speed, with devices displaying very high ~100 μs thermal time constants [66]), but it can provide a larger tuning range of about 6.6 nm [67] and induces lower power loss than current injection tuning. It is possible to only rely on one of these methods in a design, but intelligently utilizing both can enable better energy efficiency. Devices such as MRRs often use current injection tuning to switch between resonance modes and also to compensate for resonance drifts. Current injection tuning involves applying a positive/negative voltage bias to this MRR PN-junction (between the core and cladding) to inject/remove free carriers into/from the MRR core. For high frequency operation and lower power consumption, an MRR's PN-junction is typically operated under a negative voltage bias or reverse bias [68] (otherwise known as carrier depletion mode of an MRR). The application of this voltage bias generates an electric field across the MRR's core and cladding boundary. Similar to MOSFETs, this electric field generates voltage bias temperature induced (VBTI) traps at the core/cladding (Si/SiO_2) boundary of the MRR over time (*i.e.*, VBTI aging). In [69], it was shown for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

the first time that these VBTI aging induced traps alter carrier concentration in the Si core of MRRs, which incur resonance wavelength drifts and increase optical scattering loss in MRRs to degrade their quality (Q) factors. Simulation-based analysis in [69] into the impact of these changes at the system-level for the Corona and Clos photonic network-on-chip (PNoC) architectures showed that aging effects cause a worst case signal loss increase by up to 7.6 dB and energy-delay product (EDP) increase by up to 26.8%. In [96], it was shown that the use of PAM-4 signaling can reduce the impact of aging, improving energy-efficiency by 5.5% compared to using conventional on-off keying (OOK) signaling, in the presence of aging-induced long term variations.

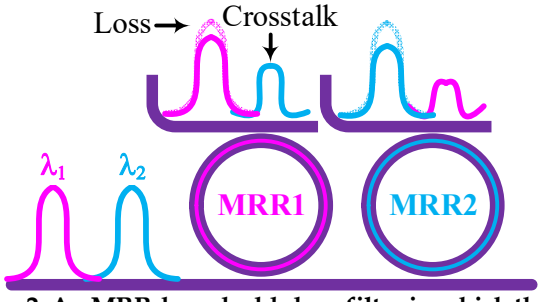


Figure 2: An MRR-based add-drop filter in which the optical signal on the wavelength λ_1 is dropped into MRR1 with some loss on the drop port while some part of the optical signal on the wavelength λ_2 (*i.e.*, unwanted crosstalk) also has been dropped into MRR1. The signals dropped into MRR2 can be similarly explained.

4 Circuit Challenges and Enhancements

Silicon photonic integrated circuits (PICs) are emerging in manycore computing systems with a promise of realizing better energy efficiency, higher bandwidth, and lower latency in such systems. Some recent notable advances in this area include a 400G silicon photonics transceiver demonstrated at Intel [16], an Intel FPGA package with an Intel Stratix 10 die integrating a silicon photonics chiplet developed at Ayar Labs targeting radar applications [18], and perhaps, artificial neural networks enabled by silicon photonics [72]. This section reviews some of the fundamental challenges at the circuit-level in terms of power loss, crosstalk, and vulnerability to different variations, all of which greatly impact the energy efficiency of silicon photonic circuits, and some state-of-the-art solutions proposed to overcome such challenges.

Signal Loss and Crosstalk Reliability

Power loss and crosstalk noise are intrinsic characteristics of fundamental silicon photonic building blocks in PICs. We discussed, for example, the propagation loss in photonic waveguides in Section 2. Looking at other sources of power loss, and in general, whenever an optical signal passes through a silicon photonic device, it suffers from some power loss and some crosstalk will be generated as well. For example, as shown in Fig. 1d, whenever an optical signal passes (OFF in Fig. 1d) or drops (ON in Fig. 1d) into an MRR, it suffers from some power loss usually known as, respectively, passing and drop loss of an MRR. In particular, crosstalk noise is of critical concern in dense wavelength-division multiplexed (WDM) circuits, where multiple optical channels exist with a small (*e.g.*, <1 nm) channel spacing. In such WDM circuits, while optical signals in each channel suffer from some optical loss, there can be intra- and inter-channel crosstalk accumulating on optical signals through different stages in the circuit (*e.g.*, through switching elements). For example, Fig. 2 indicates an MRR-based filter in which the signal on the drop port is attenuated due to power loss, and also some inter-channel crosstalk has been dropped into the MRR. Indeed, the power loss and crosstalk noise from a single silicon photonic device (*e.g.*, MRR) can be very small, and hence negligible [73]. However, in PICs integrating a large number of such devices, the small power loss and crosstalk noise at the device-level accumulate to a point that they can severely damage the performance and energy efficiency in such circuits. Some efforts have formally characterized the impact of the worst-case and average power loss and crosstalk in different PICs [74]-[80], indicating significant reductions in the optical signal-to-noise ratio (OSNR) in PICs due to high power loss and crosstalk in such circuits. Consequently, the power penalty due to crosstalk can be significant in PICs. Moreover, crosstalk increases with the data rate in PICs. For example, [81] has indicated a power penalty (due to crosstalk) of 4 dB to achieve a bit error rate of 10^{-12} at the data rate of 20 Gb/s. This power penalty increases to ~20 dB as the data rate increases to 35 Gb/s [81]. To compensate for the impact of power loss and crosstalk in PICs, one needs to increase the input laser power that considerably impact the overall energy efficiency in PICs.

Crosstalk Mitigation

Fortunately, some recent efforts have proposed solutions to reduce the impact of crosstalk in PICs. In [82], two data encoding techniques were proposed to reduce heterodyne crosstalk (*i.e.*, crosstalk noise power for a wavelength that is affected by the noise power of one or more different wavelengths) on WDM links used in large-scale PICs. In [83], the HYDRA framework was proposed that combined a different encoding technique with the use of double MRRs, and additional MRRs to mitigate inter-channel (heterodyne) crosstalk in PICs. HYDRA was shown to reduce the worst-case OSNR by up to 5.3× in PICs used across multiple PNoC architectures, but at the cost of up to 22% higher energy delay product. In [84], homodyne or intra-channel crosstalk (*i.e.*, crosstalk noise power of a wavelength that affects the signal power of the same wavelength) was modeled and a solution based on the use of a tunable decoupling waveguide was proposed to reduce this crosstalk. The solution was shown to reduce worst case OSNR by up to 37.6% in PICs used in various PNoC architectures, but at the cost of 19.2% energy overhead. These results point to the challenge with mitigating various types of crosstalk in silicon photonics: solutions to reduce crosstalk are essential for error-free communication, but their

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

use entails energy overheads that cannot be ignored [84].

Energy Minimization

Beyond crosstalk mitigation, there are a few efforts that propose solutions to minimize energy for PICs. In [85], an efficient implementation of 4-PAM signaling is presented. Compared to conventional on-off- keying (OOK) signaling, the proposed approach and implementation was able to not only reduce bit error rate (BER) by 1.5× but also reduce power by 16.9%, EPB by 14.6%, and photonic area by 10.6%. In [86], a solution was proposed to reduce laser power overheads. The proposed approach used on-chip semiconductor amplifiers (SOA) for traffic-independent and loss-aware savings in laser power. For transmitting a packet between source and destination nodes, the approach first allocates only the minimum amount of laser power to the source node that is enough for correct detection at the destination node. It then accounts for losses to be faced by the data flit on its path from the source to the destination and enables the source to amplify the allocated laser power to the necessary level by using an on-chip SOA. The approach was shown to achieve 31.5% more laser power savings with 12.8% less latency overhead compared to the best known prior work on laser power management, across PICs in PNoC architectures. Some recent efforts have also developed a power-loss-aware path mapping in PICs using lookup tables, but at a cost of memory consumption for lookup tables, which scales with the size of the circuit [87], [88].

Overcoming Impact of Variations

The fundamental silicon photonic building blocks in PICs, as mentioned in Section 2, are also vulnerable to design-time and runtime variations. Such variations result in extra power losses and crosstalk noise in devices, the impact of which rapidly accumulates in large-scale PICs, degrading the energy efficiency in such circuits. As discussed earlier, the impact of such variations can be compensated for through using circuit-level solutions (*e.g.*, active tuning) at runtime. Some efforts have been proposed to alleviate the impact of fabrication process and thermal variations at the circuit level through optical channel remapping [89], [90], intra-channel wavelength tuning and variation-aware routing [91], balanced homodyne locking [92], and using redundant devices [93]. Although these methods can help reduce the energy costs associated with circuit-level tuning, they still rely on runtime tuning of defective silicon photonic devices. Consequently, their energy cost can be in the order of mW/nm to correct a single faulty device, while the correction range (*i.e.*, nm in mW/nm) has been often minimized [89]-[91]. In terms of thermal effect mitigation, [94] and [95] developed heater proportional-integral-derivative (PID) controllers for stabilization to thermal variations.

5 Architecture Enhancements

Intra-Chip Photonic Communication

By the late 2000s, researchers began to investigate the use of photonics in chip-scale communication architectures, to improve bandwidth, latency, and energy for data movement. Early work explored on-chip bus-based hybrid electro-photonic communication architectures [97], [98] that leveraged high-speed ring-based photonic waveguides for global on-chip communication (e.g., between distant cores on the die) while conventional electrical hierarchical bus-based architectures supported local on-chip data transfers (e.g., between neighboring cores on the die). Subsequent efforts applied optical interconnects to enhance NoC architectures. In these photonic networks-on-chip (PNoCs), various types of photonic channels are leveraged to support data transfers, where a channel typically refers to a photonic waveguide which can have one or more wavelengths that are used to transfer data flits, via wavelength division multiplexing (WDM; discussed in Section 3). The degree of WDM (e.g., the use of 16, 32 or 64 wavelengths per channel) is architecture dependent.

All PNoC architectures utilize one or more of three types of photonic channels: 1) single-write-multiple-read (SWMR): where only a single node (core or memory) can write to the channel, while multiple nodes can receive the data; 2) multiple-write-single-read (MWSR): where multiple nodes can write to a channel but only one node can read from the channel; and 3) multiple-write-multiple-read (MWMR): where multiple nodes can read and write on the same channel. With exclusive sending channels, SWMR avoids starvation and does not need global arbitration (unlike MWSR) to handle contention, which reduces design complexity and network latency. When traffic loads on the channels are evenly distributed, SWMR and MWSR can perform well and provide high channel utilization. However, for unbalanced traffic distribution, their dedicated channels result in low utilization and contribute little to the network throughput. Increasing throughput would require over-provisioning of channels, which would increase static power (from the greater number of MRRs required). Therefore, the low channel utilization of SWMR and MWSR can result in low energy efficiency. MWMR channels have better utilization and network throughput due to channel sharing. Each node can write to or read from any channel via more transmitters/receivers and multiplexors than in SWMR and MWSR. Thus, under uneven traffic distribution, the nodes with high traffic injection rate can utilize multiple channels to improve channel usage. However, full channel sharing in MWMR requires more MRRs than in SWMR or MWSR channels, which can reduce its energy-efficiency.

PNoC architectures can be categorized into either 1) all-optical PNoC architectures that use optical interconnects only, or 2) hybrid PNoC architectures that combine optical and electrical interconnects. The photonic torus [99] was one of the earliest PNoC architectures proposed in literature. It consists of a photonic torus network connected to a topologically identical electronic control network that controls its operations and enables the exchange of short messages. However, the architecture suffers from waveguide crossing losses and high photonic layer area complexity. Moreover, the electrical packet switched network based photonic path setup and teardown incurs high latency and energy overheads. Inspired by [99], some researchers have

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

proposed similar switched PNoC architectures [100], [101], with active photonic routers [71], [79] that can dynamically route photonic signals (carrying data flits), similar to how data flits get routed in electrical packet-switched NoCs. Routing schemes for such architectures have also been proposed, *e.g.*, [102]. However, designing a low latency control network to dynamically tune MRRs across optical routers in such architectures is an extremely complex process [103]-[105].

In [106], the Corona all-optical crossbar topology was proposed, with photonic waveguides configured in a token-based MWSR configuration. The architecture has inspired many other crossbar based solutions, but has several shortcomings: high photonic layer complexity (more than a million MRRs required for implementation), lack of path diversity, and a reliance on expensive electro-photonic and photo-electronic conversions even for local transfers, which is inefficient. Similar to Corona, an all-optical network was proposed in [107], but based on the Clos topology. While less complex than the full crossbar topology, the topology still requires complex point-to-point photonic links and high radix photonic routers, and uses photonic interconnects even for transfers over short distances, which wastes power and leads to higher transfer latencies. BLOCON [108] is a bufferless implementation of the optical Clos with a scheduling algorithm and path allocation scheme for managing routing in the Clos. It provides low latency and high throughput, but also has higher ring heater and laser power compared to the optical Clos. The Firefly PNoC was proposed in [109], which uses a hierarchical crossbar NoC topology with clusters of nodes connected through local electrical networks, and photonic links overlaid for global, inter-cluster communication. The photonic waveguides in the architecture were configured in a reservation-assisted SWMR configuration.

In [110], the METEOR PNoC architecture was proposed (Fig. 3a) with a configurable ring-shaped crossbar that augments a traditional 2D all-electrical mesh NoC. The photonic waveguides in METEOR were configured as a combination of SWMR and MWMR (Fig. 3b) to achieve energy and latency reduction. Support for an adaptive PRI (Photonic Region of Influence) [111] allowed adaptive partitioning of traffic between the photonic crossbar and electrical 2D mesh networks. Fig. 3a shows how PRIs of different sizes can co-exist (and be reconfigured on a per-application basis), where only the cores within the PRI regions are allowed to use the photonic crossbar for communication to other cores outside their PRI. Several other efforts have proposed similar high-radix low diameter crossbar based PNoCs (*e.g.*, [112]). A few low-radix and high-diameter crossbar architectures have also been proposed [113], [114], [115]. Multi-layer PNoC architectures that leverage multiple layers of photonic devices and waveguides to reduce waveguide crossing losses have also been explored in [116] and [117].

Methods for scalable and high performance optical path setup in PNoC architectures have also been explored in a few works. For instance, [148] proposed an optical path setup approach in which path-setup messages are sent using a flooding routing strategy to enhance the probability of finding free optical paths. A few other works [110], [116], [149], rely on a ring-based path-setup networks that are able to configure multiple optical switches simultaneously instead of sequentially, thus providing more scalable path-setup performance, at the cost of a greater number of waveguides and MRRs.

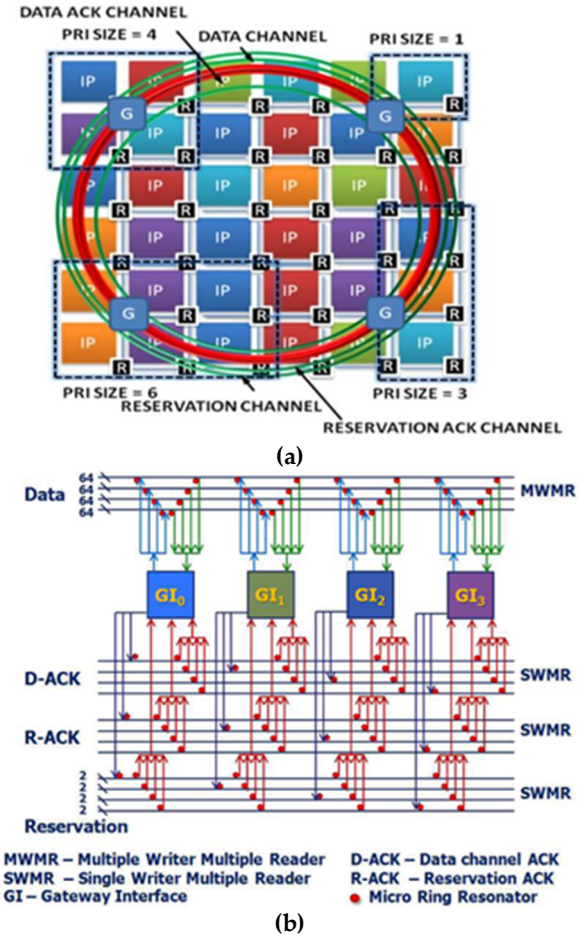
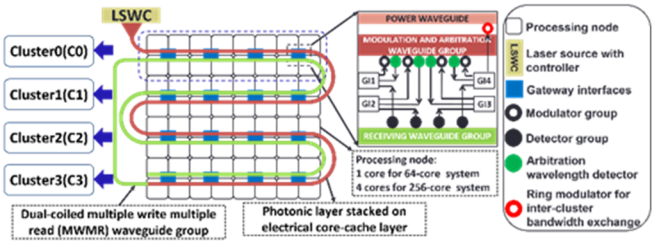


Fig. 3 (a) Meteor hybrid crossbar/mesh architecture with photonic regions of influence, (b) SWMR reservation channels and MWMR data channel configuration in METEOR [110].

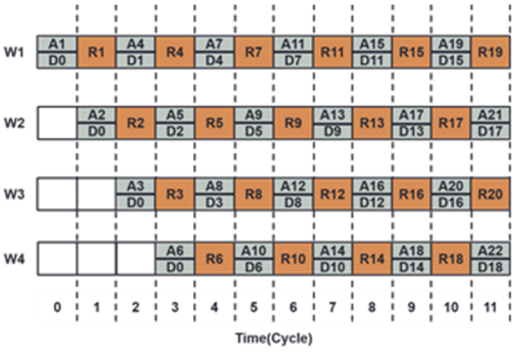
The use of MWMR or MWSR waveguides introduces the need for contention resolution protocols, to arbitrate among potentially multiple writers on the same waveguide. A few efforts have explored improved arbitration techniques that rely on time division multiplexing (TDM), so that a single data waveguide can be simultaneously used by more than one node in different time slots [118]-[121]. For instance, in Flexishare [118], a token stream arbitration scheme is proposed. The scheme requires wavelengths corresponding to each data waveguide to be injected serially into different time slots of an arbitration waveguide. A node writes on the data waveguide only when it gets access to the corresponding arbitration wavelength. Subsequently, the node cannot send data again till its arbitration wavelength is injected into the arbitration waveguide, which takes N cycles for N data waveguides. The scheme leads to channel under-utilization, and performs worse as the number of nodes and waveguides in-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

crease. In [119], the token ring arbitration scheme from Corona was improved with the token channel and token-slot arbitration techniques for MWSR crossbars. Token-slot arbitration uses time division multiplexing (TDM) and improves upon token channel arbitration by dividing the arbitration waveguide into fixed-size, back-to-back slots, with destination nodes circulating tokens in one-to-one correspondence to slots. A limitation of this approach is that a fixed time gap is required between two arbitration slots to set up data for transmission, which reduces the available time slots to send data. In UltraNoC [120], an MWMR-based PNoC architecture, a more effective concurrent token stream arbitration strategy is proposed, which together with support for reconfigurable core cluster prioritization and inter-cluster bandwidth re-allocation, is shown to improve MWMR photonic channel utilization. This technique is further improved upon in SwiftNoC [121], which allows overlapping of arbitration and data slots to reduce transfer latency, and adds more efficient data multicasting capabilities, to achieve ≈1 pJ/bit average energy efficiency for on-chip transfers.



(a)



(b)

Fig. 4 (a) Layout of MWMR crossbar used in SwiftNoC with the arrangement of cores and their respective gateway interfaces. (b) Timing diagram of arbitration in SwiftNoC, which shows distribution of arbitration (Ai), receiver selection (Ri), and data slots (Di) across four MWMR waveguide groups (W1 – W4) [121].

Fig. 4a shows the topology of the SwiftNoC PNoC architecture [121]. All cores on a chip are partitioned into four clusters (C_0 – C_3) and each cluster is assigned a dedicated arbitration wavelength (λ_0 – λ_3). Each MWMR waveguide group is divided into a fixed number of time slots, based on the time taken by light to traverse the waveguide on a die. Based on geometric calculations, each pass of the MWMR waveguide was estimated to take 4 cycles at 2.5 GHz. Thus, each MWMR waveguide group is divided into 8 time slots (4 time slots for the first pass and 4 time slots for the second pass). The time slots are further classified into three types: arbitration slot, receiver selection slot, and data slot. Fig. 4b shows an example of the distribution of time slots across 4 MWMR waveguide groups, with overlapped arbitration and data slots. In the arbitration slot, the laser source wavelength power controller (LSWC) injects the arbitration wavelengths of clusters, selectively using a modulator group to dedicate the arbitration slot to a particular cluster. Each receiving node N_i is assigned a receiver selection wavelength λ_{i+4}. Thus, after a sending node grabs an arbitration wavelength in the arbitration slot, it gets access to the next receiver selection slot which initially has all the receiver selection wavelengths injected by the LSWC. In this receiver selection slot, the sending node removes all the receiver selection wavelengths except the one corresponding to its receiving node using its modulators bank. Subsequently, in the next data slot, the sending node modulates data on the 64 wavelengths (λ_4 – λ_{67}) in each waveguide group assigned for data transfer. In the receiving portion of the MWMR waveguide (second pass of dual-coiled MWMR waveguide) whenever a receiver selection slot reaches a receiving node (N_i), the receiving node only switches-on its detector corresponding to its receiver selection wavelength λ_{i+4}. Whenever a receiving node detects its receiver selection wavelength in the receiver selection slot, it switches-on its remaining detectors to receive data in the next data slot. The stream of tokens (*i.e.*, stream of arbitration slots with arbitration wavelengths dedicated to a specific cluster) on concurrent slots in waveguide groups allows multiple nodes to inject packets simultaneously on the same MWMR waveguide, resulting in extremely high channel utilization of each MWMR waveguide group. This architecture was extended in BIGNOC [122], with homogeneous and heterogeneous photonic channel configurations, to accelerate big data application execution.

A few PNoC architectures have also been proposed that use free-space transfers [123]-[126]. They use dense Multiple Quantum Well (MQW) devices for electro-optic modulation, consuming less than 1 pJ/bit energy. These MQW devices can be configured either as absorption modulators or photodetectors (PDs). Most interestingly, MQW modulators do not suffer from the thermal variation challenges of MRRs and can be fabricated in various angles to achieve out-of-plane beam steering directions. Such free-space configurations can be integrated with standard CMOS fabrication processes. In [126], a comprehensive framework for free-space link mapping and PNoC synthesis was proposed. In the proposed architecture, MQW devices are fabricated on a GaAs substrate and then flip-chip bonded to the logic layer and waveguide coupled with a continuous wave external laser source. Modulated light can be directed through micro-mirrors and micro-lens to transmit data via the free-space medium.

Inter-Chip Photonic Communication

A few efforts have begun to explore photonics for chip-to-chip communication at the intra-board level [127]-[129]. For instance, [127] proposed the Arrayed Waveguide Grating Router (AWGR) based non-blocking, all-to-all, flat topology optical interconnect architecture. An AWGR is a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

passive optical cross connect, where every input port carries the same set of optical wavelengths, while each output port receives a set of wavelengths with each wavelength coming from a different input. This essentially creates a non-blocking, wavelength routed crossbar. The architecture showed improved energy-efficiency over electrical alternatives. But the architecture was evaluated in the C-band regime, which has limited compatibility with electro-optic printed circuit board (PCB) technology that typically offers a low waveguide loss figure at the O-band [130]. The European H2020 project ICT-STREAMS is currently working on realizing the AWGR-based interconnect benefits in the O-band and at data rates up to 50 Gb/s [131], by exploiting WDM. In [132], an AWGR-based 8-socket optical interconnect architecture was demonstrated with data rates up to 40 Gb/s, and a photonic link energy efficiency of 24 pJ/bit which can be reduced to ≈6 pJ/bit with layout enhancements, state-of-the-art ring modulator drivers, and use of Serialization/Deserialization (SerDes). This is a notable reduction over the 16.2 pJ/bit in Intel's (electrical) QPI that is widely used today.

Recent years have seen the proliferation of 2.5D integration, where small chiplets are integrated over an interposer to create a multi-chiplet processor on package. Such multiple chiplet based macro processors chips can have better yield (and thus lower cost) than a processor based on a single large die. In [133], three silicon-photonic network designs are proposed for low-power, high-bandwidth inter-chiplet communication: a static wavelength-routed point-to-point network, a "two-phase" arbitrated network, and a limited connectivity point-to-point network. Simulation results for a 64-die, 512-core cache-coherent macro chip indicate that the point-to-point network is over 10× more power-efficient and has the lowest design complexity compared to the other networks. In [134], an MWSR crossbar-based architecture is proposed to connect multiple chiplets, and is shown to achieve better performance than a concentrated mesh, Corona, and Firefly-based optical architectures adapted for inter-chiplet communication. In [135], a hybrid ring and all-to-all optical link based architecture is proposed to connect chiplets, with the ring being used to transmit data packets and the all-to-all links being used for control packets. The architecture is shown to outperform an electrical-based 2.5D interconnection solution.

6 Cross Layer Enhancements

Cross-layer approaches involve enhancements at one or more of the device, circuit, architecture, and system (operating system and/or middleware) layers in a cooperative manner. Such techniques can be significantly more effective than single-layer techniques in achieving holistic design goals such as energy efficiency.

Reliability Management

There have been a few efforts in recent years that have proposed cross-layer optimization techniques to enhance energy-efficiency and robustness in silicon photonics. In [83], the HYDRA cross-layer framework was proposed to minimize crosstalk in photonic interconnects, while improving energy-efficiency of data transfers. HYDRA combined multiple device-layer and circuit-layer techniques into a cross-layer framework. A device-layer approach was utilized for intermodulation (IM) crosstalk [136] mitigation by placing additional MRRs at modulating and receiving nodes to reduce IM noise. Another device-layer approach was utilized for heterodyne crosstalk mitigation that used double MRRs to improve worst-case OSNR in detectors by tailoring the MRRs' passbands to have steeper roll-off. Lastly, a circuit-layer technique was proposed for heterodyne crosstalk mitigation that improved worst-case OSNR in detectors by encoding data to avoid undesirable data value occurrences. The synergistic combined effect of using the two device layer and one circuit layer enhancement in HYDRA was shown to improve worst-case OSNR by up to 5.3× for the Corona and Firefly crossbar PNoC architectures in the presence of fabrication process variations, while also reducing the energy-delay-product over the best known single-layer solutions from prior work.

Variation Management

Runtime variations due to temperature changes on a chip also create a significant challenge for silicon photonics designers. For example, the resonant wavelength of an MRR is sensitive to thermal variations with up to a 7.4 nm shift in the resonance on a state-of-the-art 64-core processor chip. Such a resonance shift causes wavelength coupling (*i.e.*, inter-device matching) failures, prompting the need for dynamic MRR tuning. However, device-level tuning incurs costs in power and latency, as discussed earlier. In [150] an approach to reduce MRR tuning power via adaptive workload (thread) allocation was proposed. In [67], the LIBRA cross-layer framework was proposed that built on this idea to reduce the overhead of single-layer (device-level) optimization for overcoming the effect of thermal variations. A device-level heater proportional-integral-derivative (PID) controller was devised for stabilizing the operation of MRR devices in the presence of thermal (and process) variations. This device-layer approach was coupled with an intelligent system-layer software thread migration strategy that migrated threads to cores in a manner that reduced the energy costs of device-level thermal-induced tuning. This cross-layer approach reduced total power dissipation by up to 61% and total energy by up to 57% on the Corona and Flexishare PNoC architectures, compared to well-known single-layer optimization approaches at the device and circuit levels.

7 Open Challenges

Improving energy-efficiency in manycore computing systems with silicon photonics has received much attention over the past decade, as outlined in this survey. In this section we elaborate on some open challenges that must be addressed in the near future, that can provide opportunities for further and more aggressive energy reduction in manycore computing.

Packaging: Photonic packaging of silicon photonic integrated circuits is considerably more challenging than electronic packaging while it is also orders of magnitude more expensive. In particular, it requires robust micron-level

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/MDAT.2020.2982628, IEEE Design and Test

alignment of optical components, precise real-time temperature control, and often a high degree of vertical and horizontal electrical integration [87]. Silicon photonic packaging could be the most significant bottleneck in the development of commercially relevant integrated photonic devices at least for the next few years.

Fiber Coupling: Coupling light from an off-chip laser source into the chip always imposes some optical loss. Indeed, this coupling loss accounts for a significant source of power loss in silicon photonics [146]. Two common coupling solutions are through using surface coupling (*e.g.*, using vertical grating couplers) and edge coupling. Compared to edge coupling solutions, surface grating couplers enable wafer-scale testing and are cost-effective in terms of the fabrication process. However, their optical bandwidth is limited, and their coupling efficiency is lower. Realizing a low-loss, cost-effective, and high-bandwidth coupling solution for silicon photonic integrated circuits is one of the major challenges in this area.

Light-Source Integration: As discussed in Section 2, one of the fundamental challenges in silicon photonics to date is the lack of energy efficient on-chip lasers [19]. While off-chip lasers have high light-emitting efficiency and good temperature stability, they impose the use of inefficient, lossy couplers (discussed above) and increase packaging costs. On-chip lasers can potentially achieve a higher integration density and a better performance in terms of energy efficiency, but their development requires integration of other materials with silicon because of the low emission efficiency of silicon. Moreover, on-chip lasers must also address challenges with on-chip thermal variations, which can significantly alter their efficiencies.

Thermal and Process Sensitivity: As discussed in Sections 2 and 3, fundamental silicon photonic devices are considerably sensitive to runtime thermal variations and inevitable fabrication process variations [47]. In particular, the thermal issue is of critical concern when integrating photonics with electronics in systems like manycore computing platforms, where heat generated from electronics highly impacts photonics device and circuit performance. There are also self-heating effects in silicon photonic devices that further contributes to the thermal sensitivity problem. In terms of fabrication process variations, there are fundamental differences between variations in a conventional CMOS process and those in SP fabrication processes. Nevertheless, variation analysis and tools to enable variation-aware design in PICs are still missing and highly required. Consequently, design for manufacturability and efficient cross-layer solutions to mitigate the impact of thermal and fabrication process variations are necessary.

Electronic-Photonic Co-Design and Co-Simulation: There are multiple challenges associated with electronic-photonic co-design and co-simulation, including, but not limited to: complex nature of optical fields with both phase and amplitude, large bandwidth of optical signals that requires simulations in very small step sizes, compute-intensive optical simulations that are often required for accurate characterization of photonic devices, the need for both frequency and time domain simulations, lack of standardized behavioral models for electro-optic integration and co-simulations, lack of reliable compact models, etc. The authors refer the reader to [137] for more discussions on silicon photonics design challenges.

Design and Verification Tools: Unlike electronics, where design, simulation, test, and verification aspects are often integrated into electronic design automation (EDA) tools, tools in silicon photonics are still in an early stage [137][151]. One of the current active research areas is to develop design automation solutions, similar to those in electronics, for silicon photonics (electronic-photonic design automation). It is predictable that further improvements in such tools will facilitate the cross-layer design and system integration of silicon photonics into manycore computing platforms.

PNoC Integration: The fabrication of a manycore processor with a PNoC has still to be achieved. The major challenges are related to the number of photonic elements that can be integrated into a photonic layer. The form factors of different photonics components and design rules may create challenges to implement certain resource-hungry PNoC architectures which require tens to hundreds of thousands of devices. However, monolithic 3D and multi-layer PNoCs may be able to overcome some of these challenges. Yield at such integration scales is still a big unknown and improving the yield of electro-optic chips may necessitate new architectural innovations.

Photonic memory: Some recent efforts have begun to investigate photonic static RAM based caches [138], [139]. Chiplets with such photonic caches can avoid electro-optic conversions during data reads/writes, which can reduce the energy footprint for both data access and movement. Optical SRAM cell architectures have been demonstrated with various SOA-based layouts [140]. But building an ultra-fast optical cache memory with the capacity and energy consumption characteristics required to outperform electronic SRAM architectures is a very challenging task.

Photonic computation: There is growing interest in using photonics devices as replacements for electronics components such as transistors, to achieve computing with photons instead of electrons. Devices such as Mach-Zehnder Interferometers (MZIs), MRRs, and directional couplers allow combining light signals to emulate logic gates and perform operations such as matrix multiplications. Such photonic logic based computation promises to significantly reduce switching power (as experienced in electronic circuits). However, energy-efficient optical computing is still in its infancy, and there are many open challenges, related to scalability to large designs, area footprint reduction, power loss mitigation, and variation resilience.

QoS tradeoffs: The area of approximate computing has received a lot of interest over the past two decades. The approximate (or inexact) computing paradigm enables trading-off application output quality of service (QoS) with performance and energy goals. By sacrificing a small amount of QoS in certain application domains such as machine learning, data analytics, image processing, and database query processing, it is possible to improve performance and reduce energy consumption. This paradigm can be exploited in the photonics realm as well. The work

in [153] is one of the first efforts to enable such a trade-off and quantify benefits of approximate photonic communications with a PNoC in manycore computing system.

Security: Security represents an emerging challenge in silicon photonics. MRRs are especially susceptible to security threatening manipulations from Hardware Trojans (HTs). An HT can manipulate the tuning circuits of detector MRRs to partially tune the detector MRR to a passing wavelength in the waveguide, which enables snooping of the data that is modulated on the passing wavelength. Such covert data snooping is a serious security risk in PNoCs. SOTERIA [141] represents one of the first solutions to this challenge, but incurs non-negligible energy overheads. New solutions are needed that can improve security in energy-efficient ways.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) under grant number CCF-1813370.

REFERENCES

[1] Symposium on High Performance Interconnects, [Online]: http://www.hoti.org/

[2] International Symposium on Networks-on-Chip (NOCS), [Online]: https://www.engr.colostate.edu/nocs2019/

[3] AMD EPYC Processor Family, [Online]: https://www.amd.com/en/products/epyc

[4] Intel Xeon Platinum Processor Family, [Online]: https://www.intel.com/content/www/us/en/products/processors/xeon/scalable/platinum-processors/platinum-8180.html

[5] NVIDIA Turing Architecture, [Online]: https://www.nvidia.com/en-us/geforce/turing/

[6] AMD RDNA GPU Architecture, [Online]: https://www.amd.com/en/technologies/rdna

[7] Cerebras Wafer-Scale Engine, [Online]: https://www.cerebras.net/

[8] AMD Infinity Fabric, [Online]: https://www.amd.com/en/technologies/infinity-architecture

[9] Intel Ultra Path Interconnect Fabric, [Online]: https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview

[10] Exascale Computing Project, [Online]: https://www.exascaleproject.org/

[11] B. Dally, "Challenges for future computing systems," *in Proc. HiPEAC, Amsterdam, The Netherlands, 2015.*

[12] S. Pasricha, and N. Dutt. "On-Chip Communication Architectures", *Morgan Kauffman, ISBN 978-0-12-373892-9, Apr 2008*

[13] P. Rigby, "Three decades of innovation", *Lightwave, 31(1), 6–10, 2014.*

[14] J.W. Goodman, F. J. Leonberger, S.-Y.Kung, and R. A. Athale, "Optical interconnections for VLSI systems," *Proc. IEEE, vol. 72, pp. 850–866, 1984.*

[15] M. Lipson, "Guiding, modulating, and emitting light on Silicon: challenges and opportunities," *J. Lightw. Technol., vol. 23, no. 12, pp. 4222–4238, Dec. 2005.*

[16] Intel Silicon Photonics 100G PSM4 Optical Transceiver, [Online]: https://www.intel.com/content/www/us/en/architecture-and-technology/silicon-photonics/optical-transceiver-100gpsm4-qsfp28-brief.html

[17] Luxtera 2 × 100G-PSM4 OptoPHY Product Family. [Online]: http://www.luxtera.com/embedded-optics/

[18] Ayar Labs to Demo Photonics Chiplet in FPGA Package at Hot Chips. [Online]. Available: https://www.hpcwire.com/2019/08/19/ayar-labs-to-demo-photonics-chiplet-in-fpga-package-at-hot-chips/.

[19] Z. Zhou, B. Yin, and J. Michel, "On-chip light sources for silicon photonics," *Nature, Light: Science & Applications, vol. 4, issue 11, article e358, 2015.*

[20] D. Liang and J. Bowers, "Recent progress in lasers on silicon," *Nature Photonics, vol. 4, pp. 511–517, 2010.*

[21] K. Tanabe, K. Watanabe, and Y. Arakawa, "III-V/Si hybrid photonic devices by direct fusion bonding," *Scientific reports, vol. 2, pp. 349, 2012.*

[22] C. Zhang, D. Liang, G. Kurczveil, A. Descos and R. G. Beausoleil, "Hybrid quantum-dot microring laser on silicon," *Optica, vol. 6, pp. 1145-1151, 2019.*

[23] L. Chrostowski and M. Hochberg, "Silicon photonics Design from devices to systems," *Cambridge Univ. Press, May 2015.*

[24] M. A. Tran, D. Huang, T. Komljenovic, J. Peters, A. Malik and J. Bowers, "Ultra-low-loss silicon waveguides for heterogeneously integrated silicon/III-V photonics," *Applied Sciences, vol. 8, no. 7, article no. 1139, 2018.*

[25] I. Thakkar, S. V. R. Chittamuru, S. Pasricha, "A Comparative Analysis of Front-End and Back-End Compatible Silicon Photonic On-Chip Interconnects," *ACM/IEEE System Level Interconnect Prediction Workshop (SLIP), Jun 2016.*

[26] J. F. Bauters, M. J. R. Heck, D. John, D. Dai, M. C. Tien, J. S. Barton, A. Leinse, R. G. Heideman, D. J. Blumenthal and John E. Bowers, "Ultra-low-loss high-aspect-ratio Si3N4 waveguides," *Optics Express, vol. 19, no. 4, pp. 3163-3174, 2011.*

[27] D. Dai, Z. Wang, J. F. Bauters, M. C. Tien, M. J. R. Heck, D. J. Blumenthal and J. Bowers, "Low-loss Si3N4 arrayed-waveguide grating (de)multiplexer using nano-core optical waveguides," *Optics Express, vol. 19, no. 15, pp. 14130-14136, 2011.*

[28] R. Baets, A. Z. Subramanian, S. Clemmen, B. Kuyken, P. Bienstman, N. Le Thomas, G. Roelkens, D. Van Thourhout, P. Helin and S. Severi, "Silicon photonics: Silicon nitride versus silicon-on-insulator," *IEEE/OSA Optical Fiber Communications Conference and Exhibition (OFC), Anaheim, CA, 2016, p. Th3J.1.*

[29] J. Witzens, "High-Speed Silicon Photonics Modulators," *in Proceedings of the IEEE, vol. 106, no. 12, pp. 2158-2182, Dec. 2018.*

[30] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin,

O. Cohen, R. Nicolaescu and M. Paniccia, "'A high-speed silicon optical modulator based on a metal–oxide–semiconductor capacitor," *Nature, vol. 427, pp. 615–618, Feb. 2004.*

[31] D. Patel, A. Samani, V. Veerasubramanian, S. Gosh, and D. V. Plant, "Silicon photonic segmented modulator-based electro-optic DAC for 100 Gb/s PAM-4 generation," *IEEE Photon. Technol. Lett., vol. 27, no. 23, pp. 2433–2436, Dec. 1, 2015.*

[32] L. Alloatti, R. Palmer, S. Diebold, K. P. Pahl, B. Chen, R. Dinu, M. Fournier, J. M. Fedeli, T. Zwick, W. Freude, C. Koos and J. Leuthold, "100 GHz silicon–organic hybrid modulator," *Light, Sci. Appl., vol. 3, Mar. 2014, Art. no. e173.*

[33] M. Bahadori, M. Nikdast, Q. Cheng, and K. Bergman, "Universal design of waveguide bends in silicon-on-insulator photonics platform," in IEEE Journal of Lightwave Technology, vol. 37, no. 13, pp. 3044-3054, 2019.

[34] R. Amin, R. Maiti, C. Carfano, M. Zhizhen, M. H. Tahersima, Y. Lilach, D. Ratnayake, H. Dalir and V. J. Sorger, "0.52 V mm ITO-based Mach-Zehnder modulator in silicon photonics," *Applied Photonics, vol. 3, no. 12, Art. no. 126104, 2018.*

[35] J. Hong, F. Qiu, X. Cheng, A. M. Spring and S. Yokoyama, "A high-speed electro-optic triple-microring resonator modulator," *Scientific Reports, vol. 7, Art. no. 4682, 2017.*

[36] R. Palmer et al., "High-Speed, Low Drive-Voltage Silicon-Organic Hybrid Modulator Based on a Binary-Chromophore Electro-Optic Material," *in Journal of Lightwave Technology, vol. 32, no. 16, pp. 2726-2734, 15 Aug.15, 2014.*

[37] X. Wu, J. Xu, Y. Ye, X. Wang, M. Nikdast, Z. Wang, and Zh. Wang, "An inter/intra-chip optical network for manycore processors," *in IEEE Transactions on Very Large Scale Integration Systems, vol.23, no. 4, pp. 678-691, April 2015.*

[38] M. Briere et al., "System Level Assessment of an Optical NoC in an MPSoC Platform," *in Proc. of Design, Automation & Test in Europe Conference & Exhibition, Nice, 2007, pp. 1-6.*

[39] B. G. Lee and N. Dupuis, "Silicon Photonic Switch Fabrics: Technology and Architecture," *in Journal of Lightwave Technology, vol. 37, no. 1, pp. 6-20, 1 Jan.1, 2019.*

[40] T. J. Seok, K. Kwon, J. Henriksson, J. Luo and M. C. Wu, "240×240 Wafer-Scale Silicon Photonic Switches," *in Optical Fiber Communication Conference (OFC), OSA Technical Digest (Optical Society of America), paper Th1E.5, 2019.*

[41] P. Molly and J. E. Bowers, "Photodetectors for silicon photonic integrated circuits," *Photodetectors: Materials, Devices and Applications, pp. 3-20, 2016.*

[42] D. Benedikovic, et al., "25 Gbps low-voltage heterostructured silicon-germanium waveguide pin photodetectors for monolithic on-chip nanophotonic architectures," *Photon. Res. vol. 7, pp. 437-444, 2019.*

[43] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "Enabling efficient tolerance analysis in silicon photonic integrated circuits," in *Proc. Progress in Electromagnetic Research Symposium (PIERS), Shanghai, China, 2016, pp. 783-783.*

[44] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "An analytical study of process variations in silicon photonic integrated circuits," in *Proc. Photonics North (PN), Quebec City, Canada, 2016, pp. 1-2.*

[45] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "Photonic integrated circuits: a study on process variations," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC), Anaheim, USA, 2016, paper W2A.22.*

[46] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-ladouceur, "Silicon photonic integrated circuits under process variations," in *Proc. Asia Communications and Photonics Conference (ACP), Hong Kong, 2015, paper ASu2A.12.*

[47] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "Chip-scale silicon photonic interconnects: a formal study on fabrication non-uniformity," *IEEE Journal of Lightwave Technology, vol. 34, no. 16, pp. 3682-3695, August 2016.*

[48] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "Modeling fabrication non-uniformity in chip-scale silicon photonic interconnects," in *Proc. IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition, Dresden, Germany, 2016, pp. 115-120.*

[49] M. Nikdast, G. Nicolescu, J. Trajkovic, and O. Liboiron-Ladouceur, "DeEPeR: enhancing performance and reliability in chip-scale optical interconnection networks," in *Proc. IEEE/ACM Great Lakes Symposium on VLSI (GLSVLSI) Conference, Chicago, IL 2018, pp. 63-68.*

[50] M. Nikdast, G. Nicolescu, and O. Liboiron-Ladouceur, "Improving microresonator reliability in silicon photonic integrated circuits," in *Proc. IEEE Optical Interconnect (OI) Conference, Santa Fe, NM 2018, pp. 3-4.*

[51] A. Mirza, F. Sunny, S. Pasricha, and M. Nikdast, "Silicon photonic microring resonators: Design optimization under fabrication non-uniformity, *IEEE/ACM Design, Automation and Test (DATE) Conference and Exhibition, March 2020.*

[52] Y. Luo et al., "A process-tolerant ring modulator based on multi-mode waveguides," *IEEE Photon. Technol. Lett., vol. 28, no. 13, pp. 1391–1394, 2016.*

[53] Z. Su et al., "Reduced wafer-scale frequency variation in adiabatic microring resonators," *in OFC, 2014, p. Th2A.55.*

[54] K. Skadron et al., "Temperature-aware microarchitecture: Modeling and implementation," *ACM Trans. Archit. Code Optim., vol. 1, no. 1, pp. 94–125, 2004.*

[55] F. G. Della Corte, M. Esposito Montefusco, L. Moretti, I. Rendina, and G. Cocorullo, "Temperature dependence analysis of the thermo-optic effect in silicon by single and double oscillator models," *J. Appl. Phys., vol. 88, no. 12, pp. 7115–7119, Dec. 2000.*

[56] Y. Ye, Z. Wang, J. Xu, X. Wu, X. Wang, M. Nikdast, Zh. Wang, and L. H. K. Duong, "System-level modeling and analysis of thermal effects in WDM-based optical networks-on-chip," *in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), vol.33, no. 11, pp. 1718-1731, November 2014.*

[57] Y. Ye, J. Xu, X. Wu, W. Zhang, X. Wang, M. Nikdast, Z. Wang, and W. Liu, "System-level modeling and analysis of thermal effects in optical networks-on-chip," *in IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol. 21, no. 2, pp. 292-305, February 2013.*

[58] Y. Ye, J. Xu, X. Wu, W. Zhang, W. Liu, M. Nikdast, X. Wang, Z. Wang, and Zh. Wang, "Thermal analysis for 3D optical network-on-chip based on a novel low-cost 6×6 optical router," *Optical Interconnects Conference (OI), Santa Fe, USA, 2012, pp. 110-111.*

[59] Y. Ye, J. Xu, X. Wu, W. Zhang, X. Wang, M. Nikdast, Z. Wang, and W. Liu, "Modeling and analysis of thermal effects in optical networks-on-chip," *IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Chennai, India, 2011, pp. 254-259.*

[60] J. Ptasinski, I. C. Khoo, and Y. Fainman, "Passive Temperature Stabilization of Silicon Photonic Devices Using Liquid Crystals," *Materials (Basel), vol. 7, no. 3, pp. 2229-2241, 2014.*

[61] S. Namnabat, K. J. Kim, A. Jones, R. Himmelhuber, C. T. DeRose, D. C. Trotter, A. L. Starbuck, A. Pomerene, A. L. Lentine, and R. A. Norwood, "Athermal silicon optical add-drop multiplexers based on thermo-optic coefficient tuning of sol-gel material," *Opt. Express, vol. 25, pp. 21471-21482. 2017.*

[62] J. M. Lee, "Athermal Silicon Photonics", *In: Pavesi L., Lockwood D. (eds) Silicon Photonics III. Topics in Applied Physics, vol 122. Springer, Berlin, Heidelberg, 2016.*

[63] M. Mohamed, Z. Li, X. Chen, L. Shang, and A. R. Mickelson, "Reliability-Aware Design Flow for Silicon Photonics On-Chip Interconnect," *in IEEE TVLSI, vol. 22, no. 8, pp. 1763-1776, 2014.*

[64] J. Ahn et al., "Devices and architectures for photonic chip-scale integration," *Appl. Phys. A, vol. 95, no. 4, pp. 989–997, 2009.*

[65] M. Bahadori et al., "Thermal Rectification of Integrated Microheaters for Microring Resonators in Silicon Photonics Platform," *in Journal of Lightwave Technology, vol. 36, no. 3, pp. 773-788, 1 Feb.1, 2018.*

[66] P. Dong, W. Qian, H. Liang, R. Shafiiha, D. Feng, G. Li, J. E. Cunningham, A. V. Krishnamoorthy, and M. Asghari, "Thermally tunable silicon racetrack resonators with ultralow tuning power," *Opt. Express, vol. 18, pp. 20298-20304, 2010.*

[67] S. V. R. Chittamuru, I. Thakkar, S. Pasricha, "LIBRA: Thermal and Process Variation Aware Reliability Management in Photonic Networks-on-Chip", *IEEE Transactions on Multi-Scale Computing Systems (IEEE TMSCS), Vol. 4, No. 4, Oct-Dec 2018.*

[68] P. Dong, et al., "Low Vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator," *in Optics Express, 17:22484–22490, 2009.*

[69] S. V. R. Chittamuru, I. Thakkar, S. Pasricha, "Analyzing Voltage Bias and Temperature Induced Aging Effects in Photonic Interconnects for Manycore Computing," *ACM System Level Interconnect Prediction Workshop (SLIP), Jun 2017.*

[70] R. E. Camacho-Aguilera, Y. Cai, N. Patel, J. T. Bessette, M. Romagnoli, L. C. Kimerling, and J. Michel, "An electrically pumped germanium laser," *Optics Express, vol. 20, pp. 11316-11320, 2012.*

[71] Y. Ye, X. Wu, J. Xu, W. Zhang, M. Nikdast and X. Wang, "Holistic comparison of optical routers for chip multiprocessors," *Anti-counterfeiting, Security, and Identification, Taipei, 2012, pp. 1-5.*

[72] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," *Light: Science and Applications, vol. 8, Article no. 42, 2019.*

[73] H. Jayatilleka, K. Murray, M. Caverley, N. A. F. Jaeger, L. Chrostowski, and S. Shekhar, "Crosstalk in SOI microring resonator-based filters," *IEEE Journal of Lightwave Technology, vol. 34, no. 12, pp. 2886-2896, 2016.*

[74] L. H. K. Duong, Z. Wang, M. Nikdast, J. Xu, P. Yang, Zh. Wang, R. Maeda, H. Li, X. Wang, S. Le Beux, and Y. Thonnart, "Coherent and incoherent crosstalk noise analyses in inter/intra-chip optical interconnection networks," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol. 24, no. 7, pp. 2475- 2487, July 2016.*

[75] M. Nikdast, J. Xu, X. Wu, X. Wang, Z. Wang, Zh. Wang, and P. Yang, "Crosstalk noise in WDM-based optical networks-on-chip: A formal study and comparison," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol. 23, no. 11, pp. 2552-2565, November 2015.*

[76] M. Nikdast, J. Xu, L. H. K. Duong, X. Wu, Z. Wang, X. Wang, and Zh. Wang, "Fat-tree-based optical interconnection networks under crosstalk noise constraint," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol.23, no.1, pp. 156-169, January 2015.*

[77] M. Nikdast, J. Xu, X. Wu, W. Zhang, Y. Ye, X. Wang, Z. Wang, and Zh. Wang, "Systematic analysis of crosstalk noise in folded-torus-based optical networkson-chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), vol. 33, no. 3, pp. 437-450, March 2014.*

[78] Y. Xie, M. Nikdast, J. Xu, X. Wu, W. Zhang, Y. Ye, X. Wang, Z. Wang, and W. Liu, "Formal worst-case analysis of crosstalk noise in mesh-based optical networks-on-chip," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol. 21, no. 10, pp. 1823-1836, October 2013.*

[79] Y. Xie, M. Nikdast, J. Xu, W. Zhang, Q. Li, X. Wu, Y. Ye, W. Liu, and X. Wang, "Crosstalk noise and bit error rate analysis for optical networks-on-chip," *in Proc. Design Automation Conference (DAC), Anaheim, USA, 2010, pp. 657-660.*

[80] M. Nikdast, L. H. K. Duong, J. Xu, S. Le Beux, X. Wu,

Z. Wang, P. Yang, and Y. Ye, "CLAP: a crosstalk and loss analysis platform for optical interconnects," *in Proc. IEEE/ACM International Symposium on Networks-on-Chip (NoCS), Ferrara, Italy, 2014, pp. 172-173.*

[81] M. Bahadori et al., "Crosstalk Penalty in Microring-Based Silicon Photonic Interconnect Systems," *Journal of Lightwave Technology, vol. 34, no. 17, pp. 4043–4052, Sep. 2016.*

[82] S. V. R. Chittamuru, S. Pasricha, "Crosstalk Mitigation for High-Radix and Low-Diameter Photonic NoC Architectures", *IEEE Design and Test (D&T), vol.32, no.3, pp.29-39, June 2015.*

[83] S. V. R. Chittamuru, I. Thakkar, S. Pasricha, "HYDRA: Heterodyne Crosstalk Mitigation with Double Microring Resonators and Data Encoding for Photonic NoC", *IEEE Transactions on Very Large Scale Integration Systems (TVLSI), vol. 26, iss. 1, pp. 168 – 181, Jan 2018.*

[84] I. Thakkar, S. V. R. Chittamuru, S. Pasricha, "Mitigation of Homodyne Crosstalk Noise in Silicon Photonic NoC Architectures with Tunable Decoupling," *in Proc. ACM/IEEE International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS), Oct 2016.*

[85] I. Thakkar, S. V. R. Chittamuru, S. Pasricha, "Improving the Reliability and Energy-Efficiency of High-Bandwidth Photonic NoC Architectures with Multilevel Signaling," *in Proc. IEEE/ACM International Symposium on Networks-on-Chip (NOCS), Oct 2017.*

[86] I. Thakkar, S. V. R. Chittamuru, S. Pasricha, "Run-Time Laser Power Management in Photonic NoCs with On-Chip Semiconductor Optical Amplifiers," *in Proc. IEEE/ACM International Symposium on Networks-on-Chip (NOCS), Aug 2016.*

[87] Q. Cheng, M. Bahadori, Y. Huang, S. Rumley, and K. Bergman, "Smart routing tables for integrated photonic switch fabrics," *in Proc. IEEE European Conference on Optical Communication (ECOC), 2017, pp. 1–3.*

[88] Q. Cheng, M. Bahadori, and K. Bergman, "Advanced path mapping for silicon photonic switch fabrics," *in Proc. IEEE/OSA Conference on Lasers and Electro-Optics (CLEO), 2017, paper SW1O.5.*

[89] Y. Xu, J. Yang, and R. Melhem, "Tolerating process variations in nanophotonic on-chip networks," *in Proc. Annual International Symposium on Computer Architecture, Jun. 2012, pp. 142–152.*

[90] Y. Xu, J. Yang, and R. Melhem, "BandArb: Mitigating the effects of thermal and process variations in silicon-photonic network," *in Proc. ACM Int. Conf. Comput. Frontiers, May 2015, pp. 30-1–30-8.*

[91] M. Mohamed, Z. Li, X. Chen, L. Shang, A. Mickelson, M. Vachharajani, and Y. Sun, "Power-efficient variation-aware photonic on-chip network management," *in Proc. ACM/IEEE International Symposium on Low-Power Electronics and Design (ISLPED), Austin, TX, USA, 2010, pp. 31-36.*

[92] J. A. Cox, A. L. Lentine, D. C. Trotter, and A. L. Starbuck, "Control of integrated micro-resonator wavelength via balanced homodyne locking," *Optics Express, vol. 22, pp. 11279-11289. 2014.*

[93] M. Georgas, J. Leu, B. Moss, C. Sun, and V. Stojanovic, "Addressing link-level design tradeoffs for integrated photonic interconnects," *in Proc. IEEE Conference on Custom Integrated Circuits, 2011, pp. 1–8.*

[94] D. Dang, S. V. R. Chittamuru, R. N. Mahapatra, S. Pasricha, "Islands of Heaters: A Novel Thermal Management Framework for Photonic NoCs," *in Proc. IEEE/ACM Asia & South Pacific Design Automation Conference (ASPDAC), Jan 2017.*

[95] D. Dang, S. V. R. Chittamuru, R. N. Mahapatra, S. Pasricha, "Islands of Heaters: A Novel Thermal Management Framework for Photonic NoCs," *in Proc. IEEE/ACM Asia & South Pacific Design Automation Conference (ASPDAC), Jan 2017.*

[96] I. Thakkar, S. V. R. Chittamuru, S. Pasricha, "Mitigating the Energy Impacts of VBTI Aging in Photonic Networks-on-Chip Architectures with Multi-level Signaling," *in Proc. IEEE Workshop on Energy-efficient Networks of Computers (E2NC): from the Chip to the Cloud, Oct 2018.*

[97] N. Kirman et. al, "Leveraging Optical Technology in Future Bus-based Chip Multiprocessors", *In Proc. MICRO, 2006.*

[98] S. Pasricha, N. Dutt, "ORB: An On-chip Optical Ring Bus Communication Architecture for Multi-Processor Systems-on-Chip", *in Proc. IEEE Asia & South Pacific Design Automation Conference (ASPDAC), Seoul, Korea, Jan 2008.*

[99] A. Shacham, K. Bergman and L. P. Carloni, "Photonic Networks-on-Chip for Future Generations of Chip Multiprocessors," *IEEE Transactions on Computers, vol. 57, no. 9, pp. 1246-1260, Sept. 2008.*

[100] Z. Chen, H. Gu, Y. Yang, and D. Fan, "A hierarchical optical network-on-chip using central-controlled subnet and wavelength assignment," *J. Lightw. Technol., vol. 32, no. 5, pp. 930-938, Mar. 2014.*

[101] W. Tan, H. Gu, Y. Yang, K.Wang, and X. Wang, "Venus: A Low-Latency, Low-Loss 3D hybrid Network-on-Chip for Kilocore Systems," *IEEE J. Lightw. Technol., vol. PP, no. 99, pp. 1-8, 2017.*

[102] P. Guo et al., "Fault-Tolerant Routing Mechanism in 3D Optical Network-on-Chip based on Node Reuse," *IEEE Transactions on Parallel and Distributed Systems, 2019.*

[103] F. Gohring, R. Priti, M. Nikdast, F. Hessel, O. Liboiron-Ladouceur, and G. Nicolescu, "Design and modelling of a low-latency centralized controller for optical integrated networks," *IEEE Communications Letters (COMML), vol. 20, no. 3, pp. 462-465, March 2016.*

[104] F. Gohring, M. Nikdast, Y. Xiong, F. Hessel, O. Liboiron-Ladouceur, and G. Nicolescu, "Silicon photonic interconnects: minimizing the controller latency," *in Proc. ACM Great Lakes Symposium on VLSI (GLSVLSI) Conference, Chicago, IL 2018, pp. 323-328.*

[105] F. Gohring, R. Priti, M. Nikdast, F. Hessel, O. Liboiron-Ladouceur, and G. Nicolescu, "A low-latency centralized controller for MZI-based optical integrated networks," *in Proc. International Conference on Photonics in Switching (PS), Florence, Italy, 2015, pp. 118-120.*

[106] D. Vantrease et al., "Corona: System implications of emerging nanophotonic technology," *in Proc. Int. Symp. Comput. Archit., 2008, pp. 153–164.*

[107] A. Joshi, C. Batten, Y.-J. Kwon, S. Beamer, I. Shamim, K. Asanovic, and V. Stojanovic, "Silicon-photonic clos networks for global on-chip communication", *in Proc.* ACM/IEEE International Symposium on Networks-on-Chip (NOCS), 2009.

[108] Y.-H. Kao and H. J. Chao, "Blocon: A bufferless photonic clos network-on-chip architecture," *in Proc. ACM/IEEE NOCS, pp. 81–88, IEEE, 2011.*

[109] Y. Pan, P. Kumar, J. Kim, G. Memik, Y. Zhang, and A. Choudhary, "Firefly: Illuminating future network-on- chip with nanophotonics", *in Proc. International Symposium on Computer Architecture (ISCA). 429–440, 2009.*

[110] S. Bahirat, S. Pasricha, "METEOR: Hybrid Photonic Ring-Mesh Network-on-Chip for Multicore Architectures", *ACM Transactions on Embedded Computing Systems (TECS), 13(3):116:1-116:33, Mar 2014.*

[111] S. Bahirat, S. Pasricha, "UC-PHOTON: A Novel Hybrid Photonic Network-on-Chip for Multiple Use-Case Applications", *in Proc. IEEE International Symposium on Quality Electronic Design (ISQED) Santa Clara, CA, Mar 2010.*

[112] J. Psota, J. Miller, G. Kurian, H. Hoffman, N. Beckmann, J. Eastep, and A. Agarwal, "ATAC: Improving performance and programmability with on-chip optical networks", *in Proc. IEEE International Symposium on Circuits and Systems (ISCAS), 2010.*

[113] Y. H. Kao and H. J. Chao, "BLOCON: A Bufferless Photonic Clos network-on-chip architecture", *in Proc. IEEE/ACM International Symposium on NoCS, 2011.*

[114] S. Werner, J. Navaridas and M. Luján, "Designing Low-Power, Low-Latency Networks-on-Chip by Optimally Combining Electrical and Optical Links," *in Proc. IEEE International Symposium on High Performance Computer Architecture (HPCA), Austin, TX, 2017, pp. 265-276.*

[115] R. Morris and A. K. Kodi, "Exploring the design of 64-and 256-core power efficient nanophotonic interconnect," *IEEE Journal of Selected Topics in Quantum Electronics, vol. 16, pp. 1386–1393, 2010.*

[116] S. Pasricha, S. Bahirat, "OPAL: A Multi-Layer Hybrid Photonic NoC for 3D ICs", *in Proc. IEEE/ACM Asia & South Pacific Design Automation Conference (ASPDAC), Japan, Jan 2011.*

[117] X. Zhang and A. Louri, "A multilayer nanophotonic interconnection network for on-chip many-core communications", *in Proc. Design Automation Conference (DAC), 156–161, Jun 2011.*

[118] Y. Pan, J. Kim, and G. Memik, "Flexishare: Channel sharing for an energy efficient nanophotonic crossbar", *in Proc.* International Symposium on High Performance Computer Architecture (HPCA). 1-12, 2010.

[119] D. Vantrease, N. Binkert, R. Schreiber, and M. H. Lipasti. "Light speed arbitration and flow control for nanophotonic interconnects", *in Proc. IEEE/ACM International Symposium on Microarchitecture (MICRO'09), 304-315, 2009.*

[120] S. V. R. Chittamuru, S. Desai, S. Pasricha, "A Reconfigurable Silicon-Photonic Network with Improved Channel Sharing for Multicore Architectures," *in Proc. ACM GLSVLSI, May 2015.*

[121] S. V. R. Chittamuru, S. Desai, S. Pasricha, "SWIFT-NoC: A Reconfigurable Silicon-Photonic Network with Multicast Enabled Channel Sharing for Multicore Architectures", *ACM Journal on Emerging Technologies in Computing Systems (JETC), Vol. 13, No. 4, pp. 58:1-58:27, Jun 2017.*

[122] S. V. R. Chittamuru, D. Dharnidhar, S. Pasricha, R. Mahapatra "BiGNoC: Accelerating Big Data Computing with Application-Specific Photonic Network-on-Chip Architectures," *IEEE Transactions on Parallel and Distributed Systems (TPDS), Vol. 29, Iss. 11, Nov 2018.*

[123] S. Bahirat, S. Pasricha, "HELIX: Design and Synthesis of Hybrid Nanophotonic Application-Specific Network-On-Chip Architectures", *in Proc. IEEE International Symposium on Quality Electronic Design (ISQED), Mar. 2014.*

[124] P. Guo, W. Hou, L. Guo, X. Zhang, Z. Ning and M. S. Obaidat, "Design for Architecture and Router of 3D Free-Space Optical Network-on-Chip," *in Proc. IEEE International Conference on Communications (ICC), Kansas City, MO, 2018, pp. 1-6.*

[125] J. Xue, A. Garg, B. Ciftcioglu, J. Hu, S. Wang, I. Savidis, M. Jain, R. Berman, P. Liu, M. C. Huang, H. Wu, E. G. Friedman, G. Wicks, and D. Moore, "An intra-chip free-space optical interconnect", *in Proc. of ISCA, pp 94–105, 2010.*

[126] S. Bahirat, S. Pasricha, "3D HELIX: Design and Synthesis of Hybrid Nanophotonic Application-Specific 3D Network-On-Chip Architectures", *Workshop on Exploiting Silicon Photonics for Energy efficient Heterogeneous Parallel Architectures (SiPhotonics), Jan. 2014.*

[127] P. Grani, R. Proietti, S. Cheung, and S. J. B. Yoo, "Flat-topology high throughput compute node with AWGR-based optical-interconnects," *J. Lightw. Technol., vol. 34, no. 12, pp. 2959–2968, Jun. 2016.*

[128] C. Batten et al., "Building Manycore Processor-to-DRAM Networks with Monolithic CMOS Silicon Photonics," *in IEEE Micro, 2009.*

[129] P. Grani, R. Proietti, V. Akella and S. J. B. Yoo, "Design and Evaluation of AWGR-Based Photonic NoC Architectures for 2.5D Integrated High Performance Computing Systems," *in IEEE International Symposium on High Performance Computer Architecture (HPCA), Austin, TX, 2017, pp. 289-300, 2017.*

[130] A. Sugama, K.Kawaguchi, M.Nishizawa, H.Mura-

naka, and Y.Arakawa, "Development of high-density single-mode polymer waveguides with low crosstalk for chip-to-chip optical interconnection," *Opt. Express, vol. 21, no. 20, 2013, Art. no. 24231.*

[131] G. T. Kanellos and N. Pleros, "WDM mid-board optics for chip-to-chip wavelength routing interconnects in the H2020 ICT-STREAMS," *in Proc. SPIE, Opt. Interconnects XVII, vol. 10109, Feb. 2017.*

[132] S. Pitris et al., "A 40 Gb/s Chip-to-Chip Interconnect for 8-Socket Direct Connectivity Using Integrated Photonics," *in IEEE Photonics Journal, vol. 10, no. 5, pp. 1-8, Oct. 2018*

[133] P. Koka, M.McCracken, H. Schwetman, X. Zheng, R. Ho, and A. Krishnamoorthy, "Silicon-photonic network architectures for scalable, power efficient multi-chip systems," *in ACM SIGARCH Comput. Archit. News, vol. 38, no. 3, 2010, Art. no. 117.*

[134] Y. Demir et al., "Galaxy: A high-performance energy-efficient multichip architecture using photonic interconnects," *in Proc. ACM Int. Conf. Supercomput., Munich, Germany, Jun. 2014, pp. 303–312.*

[135] Z. Wang et al., "CAMON: Low-Cost Silicon Photonic Chiplet for Manycore Processors," *in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.*

[136] K. Padmaraju et al., "Intermodulation Crosstalk Characteris-tics of WDM Silicon Microring Modulators," *in PTL, 2014.*

[137] W. Bogaerts and L. Chrostowski, "Silicon Photonics Circuit Design: Methods, Tools and Challenges," *Laser & Photonics Reviews, vol. 12, no. 4, p. 1700237, Apr. 2018.*

[138] P. Maniotis, S. Gitzenis, L. Tassiulas, and N. Pleros, "An optically enabled chip-multiprocessor architecture using a single-level shared optical cache memory," *Opt. Switching Netw. J., vol. 22, pp. 54–68, Nov. 2016.*

[139] C. Vagionas et al., "All-optical tag comparison for hit/miss decision in optical cache memories" *IEEE Photon. Technol. Lett., vol. 28, no. 7, pp. 713–716, Dec. 2015.*

[140] T. Alexoudi et al., "III-V-on-Si photonic crystal nanocavity laser technology for optical static random access memories (SRAMs)," *IEEE J. Sel. Topics Quantum Electron., vol. 22, no. 6, pp. 1–10, Nov./Dec. 2016.*

[141] S. V. R. Chittamuru, I. Thakkar, S. Pasricha, "SOTERIA: Exploiting Process Variations to Enhance Hardware Security with Photonic NoC Architectures," *IEEE/ACM Design Automation Conference (DAC), San Francisco, CA, USA, Jun. 2018.*

[142] S. Matsuo, T. Fujii, K. Hasebe, K. Takeda, T. Sato, and T. Kakitsuka, "Directly modulated buried heterostructure DFB laser on SiO2/Si substrate fabricated by regrowth of InP using bonded active layer," *Opt. Express*, vol. 22, *pp. 12139-12147, 2014.*

[143] J. Sun, R. Kumar, M. Sakib, J. B. Driscoll, H. Jayatilleka and H. Rong, "A 128 Gb/s PAM4 Silicon Microring Modulator With Integrated Thermo-Optic Resonance Tuning," *IEEE Journal of Lightwave Technology, vol. 37, no. 1, pp. 110-115, 1 Jan.1, 2019.*

[144] W. Bogaerts, P. W. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators", *Laser & Photon. Rev., vol. 6, pp. 47-73, 2012.*

[145] C. Errando-Herranz, F. Niklaus, G. Stemme, and K. B. Gylfason, "Low-power microelectromechanically tunable silicon photonic ring resonator add-drop filter," *Opt. Lett., vol. 40, pp. 3556-3559, 2015.*

[146] R. Marchetti, C. Lacava, L. Carroll, K. Gradkowski, and P. Minzioni, "Coupling strategies for silicon photonics integrated chips [Invited]," *Photon. Res., vol. 7, pp. 201-239, 2019.*

[147] N. M. Fahrenkopf, C. McDonough, G. L. Leake, Z. Su, E. Timurdogan and D. D. Coolbaugh, "The AIM Photonics MPW: A Highly Accessible Cutting Edge Technology for Rapid Prototyping of Photonic Integrated Circuits," *IEEE Journal of Selected Topics in Quantum Electronics, vol. 25, no. 5, pp. 1-6, Sept.-Oct. 2019, Art no. 8201406.*

[148] E. Fusella, J. Flich, and A. Cilardo. "Path setup for hybrid NoC architectures exploiting flooding and standby" *IEEE Transactions on parallel and Distributed systems 28, no. 5 (2016): 1403-1416.*

[149] P. Grani, S. Bartolini, "Simultaneous optical path-setup for reconfigurable photonic networks in tiled CMPS," *in Proc. IEEE Int. Conf. High Performance Comput. Commun. (HPCC), 2014, pp. 482–485.*

[150] J. L. Abellán, A. K. Coskun, A. Gu, W. Jin, A. Joshi, A. B. Kahng, J. Klamkin, C. Morales, J. Recchio, V. Srinivas, T. Zhang, "Adaptive Tuning of Photonic Devices in a Photonic NoC Through Dynamic Workload Allocation" *in IEEE Trans. on CAD of Integrated Circuits and Systems 36(5): 801-814 2017.*

[151] L. Chrostowski, J. Flueckiger, C. Lin, M. Hochberg, J. Pond, J. Klein, J. Ferguson, C. Cone, "Design methodologies for silicon photonic integrated circuits," *Proc. SPIE 8989, Smart Photonic and Optoelectronic Integrated Circuits XVI,* vol. 8989, pp. 83-97, 2014.

[152] C. A. Thraskias et al., "Survey of Photonic and Plasmonic Interconnect Technologies for Intra-Datacenter and High-Performance Computing Communications," *in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2758-2783, Fourth Quarter 2018.*

[153] F. Sunny, A. Mirza, I. Thakkar, S. Pasricha, and M. Nikdast, "LORAX: Loss-Aware Approximations for Energy-Efficient Silicon Photonic Networks-on-Chip", *ACM Great Lakes Symposium on VLSI (GLSVLSI), 2020,*

Sudeep Pasricha (sudeep@colostate.edu) received his Ph.D. in computer science from the University of California, Irvine in 2008. He is currently a Professor at Colorado State University. His research interests include networks-on-chip, and hardware/software co-design for energy-efficient, secure, and fault-tolerant embedded systems. He is a Senior Member of IEEE.

Mahdi Nikdast (mnikdast@colostate.edu) received his Ph.D. in ECE from the Hong Kong University of Science and Technology in 2013. He is currently an Assistant Professor of ECE at Colorado State University. His research interests include silicon photonics, high-performance computing, and emerging hardware technologies. He is a Senior Member of IEEE.

Mail Address : 1373 Campus Delivery, Colorado State University, Fort Collins, CO 80523-1373

.