Reliability-Aware Design Flow for Silicon Photonics On-Chip Interconnect

Moustafa Mohamed, *Student Member, IEEE,* Zheng Li, *Student Member, IEEE*, Xi Chen, *Student Member, IEEE*, Li Shang, *Member, IEEE*, and Alan R. Mickelson, *Senior Member, IEEE*

***Abstract*— Intercore communication in many-core processors presently faces scalability issues similar to those that plagued intracity telecommunications in the 1960s. Optical communication promises to address these challenges now, as then, by providing low latency, high bandwidth, and low power communication. Silicon photonic devices presently are vulnerable to fabrication and temperature-induced variability. Our fabrication and measurement results indicate that such variations degrade interconnection performance and, in extreme cases, the interconnection may fail to function at all. In this paper, we propose a reliability-aware design flow to address variation-induced reliability issues. To mitigate effects of variations, limits of device design techniques are analyzed and requirements from architecture-level design are revealed. Based on this flow, a multilevel reliability management solution is proposed, which includes athermal coating at fabrication-level, voltage tuning at device-level, as well as channel hopping at architecture-level. Simulation results indicate that our solution can fully compensate variations thereby sustaining reliable on-chip optical communication with power efficiency.**

***Index Terms*— Multicore processing, multiprocessor interconnection networks, nanophotonics, reliability.**

I. Introduction

SILICON photonics provide the low latency and high bandwidth with low power dissipation that is required for the thousand-core computing [1]. Relay-free optical communication offers low latency, a stringent requirement of many-core systems. Wave Division Multiplexing (WDM) offers high bandwidth with small area overhead. The number of channels multiplexed on a waveguide is expected to scale with technology advancement. Power efficiency of a photonic point-to-point communication link is superior to its electrical counterpart [2]. Computer architects have proposed several innovative approaches to leverage optics and overcome its limitations [2]–[4].

However, silicon photonic interconnect faces serious reliability challenges. Silicon photonic devices are sensitive to process and thermal variations [5]. Typical operating and fabrication conditions will give rise to performance degradation [6], [7]. Microring measurements show that spatial and temporal thermal variations in many-core systems may result in communication link failures [7]. Process variations can further increase latency by an order of magnitude [8].

Existing reliability research has followed two separate paths. Device designers have attempted to reduce thermal sensitivity of devices without a vision of what level of sensitivity is acceptable from a system perspective [9]. The proposed techniques either do not address the variations problem to the full range or come with expensive power or fabrication cost. On the other hand, system designers ignore the problem altogether [2] or use inefficient solutions. For instance, a recent design allocates 54% of its total power for thermal tuning [4]. Recently, reliability management techniques have been proposed that make minimal assumptions about the system under study [6], [7], [10], [11]. Despite high reliability, the runtime power and performance overhead associated with these solutions is significant and does not scale to future many-core systems. Unnecessary overhead for control and compensation of thermal variations introduces additional degradation in power, performance, and scalability.

We propose, herein, a reliability-aware design flow to thoroughly analyze the reliability challenges, compare available reliability techniques, and devise a multilevel reliability management solution that meets specific design requirements and outperforms existing ones. The flow and the proposed solution are based on hundreds of basic silicon photonic devices that we have designed, fabricated, measured, and modeled. We have studied them in light of process and thermal variations and quantified their impact at the system-level.

The proposed design flow relies on reliability-aware analysis, modeling, and design of two tightly coupled abstraction levels: device and network levels. The flow starts with analysis using device-level reliability models, and the exploration of network-level design space. The outcome from both the device- and network-levels is leveraged in designing a reliability management solution that optimizes system power, performance, and reliability as a whole.

We apply the design flow on a specific system and a set of workloads, demonstrating the analysis and design steps through different case studies. The outcome is a reliability management solution for our specific application and workload that is more efficient in terms of power, performance, and scalability. Specifically, three techniques are com-

Manuscript received December 31, 2012; revised May 27, 2013; accepted July 25, 2013. Date of publication October 23, 2013; date of current version July 22, 2014. This work was supported by the NSF under Award CCF-0829950 and Award CCF-0954157.

The authors are with the University of Colorado, Boulder, CO 80309 USA (e-mail: moustafa.mohamed@colorado.edu; zheng.li@colorado.edu; xi.chen@colorado.edu; li.shang@colorado.edu; alan.mickelson@colorado.edu).

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

Digital Object Identifier 10.1109/TVLSI.2013.2278383

bined to address process and thermal variations. we propose: 1) athermal polymer coating, a fabrication-level technique, to overcome the thermal sensitivity obviating the need for thermal management techniques and reducing the run-time overhead; 2) channel hopping, a network-level solution, for coarse-grain level matching of senders with receivers of similar channel passband; and 3) voltage tuning, a device-level solution, for fine tuning the passband mismatch to eliminate the effect of process variations.

Simulation results based on device measurements show that the proposed solution provides a reliable, power-efficient, and scalable photonic interconnect for many-core systems. We compare our reliability management solution to four recently proposed reliability management techniques [6], [7], [10], [11] in a simulation-based study. Our solution consumes <2% of total network power under different workloads and consumes less power than earlier techniques as it eliminates the need for run-time control that other techniques rely on. Additionally, our solution exhibits scalability to future many-core systems, thanks to the elimination of run-time overhead. Meanwhile, previous techniques rely on the power-inefficient run-time thermal tuning technique, which makes them difficult to scale.

This paper is organized as follows: we assess the reliability problem from the device- and system-levels showing the serious system impact due to variations (Section II). In Section III, we describe reliability-aware design flow in light of our two abstraction levels: device- and network-level. In addition, we discuss the different reliability management techniques at each level. In Section IV, we propose a novel reliability management solution based on the analysis we conducted in our device- and network-level. In Section V, we present and discuss design and simulation results that evaluates our reliability management solution and compares it to existing solutions. Finally we conclude in Section VI.

II. Reliability Challenges

Process and thermal variations of silicon photonic devices are pervasive. Their impact is substantial. In this section, we quantify the device variations and model the impact they have on system reliability.

A. Variation-Induced Reliability Challenges

The silicon photonic interconnection fabric depends on the matching of resonant wavelengths of multiple, spatially separated, resonant devices used for signal modulation, switching, and filtering [2].

However, resonant device characteristics (e.g., passband wavelengths of microrings and racetracks) are sensitive to device dimensions and refractive indices of materials. When passbands of receivers and transmitters do not fully overlap, signal loss and crosstalk occurs, which leads to degradation of system performance and increased power dissipation.

Process variations are statistical. They are the variances of actual processes. The characterization requires analysis of sets of measurement data. Fabrication-induced process variations affects the critical dimensions of silicon photonic devices, most prominently the waveguide width and thickness, resulting

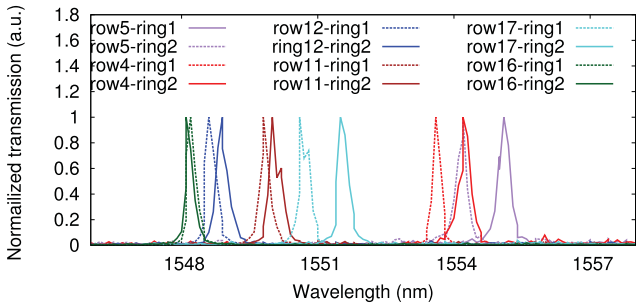

Fig. 1. Passband shift of microrings due to process variations.

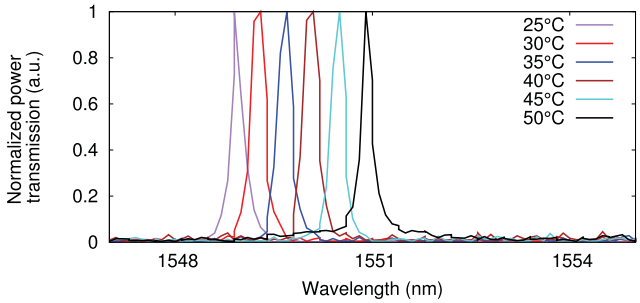

Fig. 2. Passband shift of microrings due to thermal variations.

in passband shift (our nominal waveguide dimensions are 450×220 nm^2). To quantify the process variations, we have fabricated more than six wafers at ePIXfab [12], each covered with multiple batches of devices. On one of the 200 mm wafers, we have measured and characterized over 20 dies with same nominal design dimensions. Fig. 1 displays the passbands of 12 different microrings with the same design in different locations of the wafer (from 6 different chips and 2 microrings on each chip). The microring under study has a radius of 4.98 μm and a gap of 200 nm. As shown, passband shift is so great so that channels no longer overlap at different locations on the wafer.

These variations are a result of thickness and width variations of the silicon layer that coats the oxide layer of the silicon on oxide (SoI) wafer. Selvaraja *et al.* [13] shows that the thickness variations has a standard deviation of 2 nm and a range 7.1 nm. The width variations have a standard variation of 2.59 nm and a range 7.5 nm. These results explain the significant shift in resonant wavelength shown in Fig. 1.

Thermal variations is the result of spatial and temporal variations in temperature. In practice, large (tens of celsius degrees) temperature variations may be the caused by the proximity of CMOS processors that emit varying amounts of heat with time. Previous thermal analysis shows that the spatial variations across of 1 cm^2 chip may be 17° and the temporal variations may reach 40° [7]. Thermal variations causes the passband of silicon photonic devices to drift as a result of the thermo-optic coefficient of silicon. To quantify the effect on photonic devices, we subjected samples under test to thermal variations. The microrings and racetrack resonators tested exhibited change in spectral transmission. Fig. 2 displays the dependence of the passband of a microring resonator on temperature over a 50° temperature range. The experiment

was conducted on a microring with a radius of 4.98 μm and gap 200 nm. As can be seen, the drift in the center wavelength of the passband is roughly 0.11 nm/°C. Wavelength shift for racetracks is smaller, approximately 0.09 nm/°C according to our measurements [5].

B. System Reliability Impact

To model the impact on system reliability of variation-induced device passband mismatches, we modeled the optical signal path to determine the signal-to-noise degradation at the receiver. In a typical silicon photonic interconnect, there are four separate devices that must have matched passbands: the demultiplexer at the broadband light source, the multiplexer that modulates the signals before transmission, the switch, and the demultiplexer at the receiver end. Not only is the signal degraded from loss, but noise is also introduced through crosstalk, a secondary effect of passband drift. That is, power that is not coupled in the passband may still find a way through the system to another receiver. The damage caused by thermal drift is therefore twofold: lower signal levels and higher noise levels lead to a lower signal-to-noise ratio (SNR) at the detector. This SNR is quantitatively defined by [6]

$$\mathrm{SNR} = \frac{\mathcal{R} P_{\mathrm{rec}}}{N_{\mathrm{r}} + N_{\mathrm{x}}} \tag{1}$$

where SNR is signal-to-noise ratio, $\mathcal{R}$ is photodetector responsivity, P_{rec} is signal level received at detector, N_{r} is receiver noise such as thermal noise, shot noise, and dark current noise (dominated by thermal noise), and N_{x} is crosstalk noise.

Degradation of SNR directly affect the ability of the receiver to distinguish between the on and off signals representing ones and zeros, thus impacting the reliability of communication. The reliability of a communication link is measured in terms of bit-error-rate and is defined as

$$\mathrm{BER} = 1/2 erfc(t\sqrt{\mathrm{SNR}}/\sqrt{2}) \tag{2}$$

where t is the minimum distance of the error correcting code.

Variations too small to result in catastrophic failure still result in performance degradation. When a bit error is detected, retransmission is necessary, resulting in increased latency for the packet. The packet latency, in terms of bit-error-rate, can be defined as [7]

$$t = t_o + (1-p)\, t_o + (1-p)^2\, t_o + \ldots = t_o / (1 - \mathrm{BER})^m . \tag{3}$$

Retransmission has two side effects: first, congestion in the network occurs; second, packets on the critical path of the execution directly affect the system throughput.

As one can see from (2), the bit-error-rate of the communication channel depends exponentially on the SNR. Small fluctuations in received power may lead to dramatic reliability degradation. For a bit-error-rate of 10^{-12} (that corresponds to one-bit error every 250 s per channel at 4 Gb/s) required for reliable communication, a signal-to-noise ratio of 49 is required. Measurements indicate that drifts in microring passbands may exceed the passband width. Using 0.11 nm/°C and 1.48 nm passband for a 64-channel system operating in the C+L band, a 13° spatial thermal variations is enough to totally block communication [7]. Similarly, process variations may be large enough to result in reliability failure. Our measurements of SoI microrings fabricated in ePIXfab result in a 1.2 nm standard deviation in resonant wavelength within a single die [5]. The design had a high device density, a number of gaps near the fabrication limit (130 nm), and radii as small as 5 μm. The designs in [13] had a lower device density, larger gaps of 180 nm, and minimum radii of 5, 5.01, 5.02, 5.03 μm. Selvaraja measured a 0.3–0.6 nm standard deviation in resonant wavelength shift. Our measured variations are larger than other numbers presented in [13].

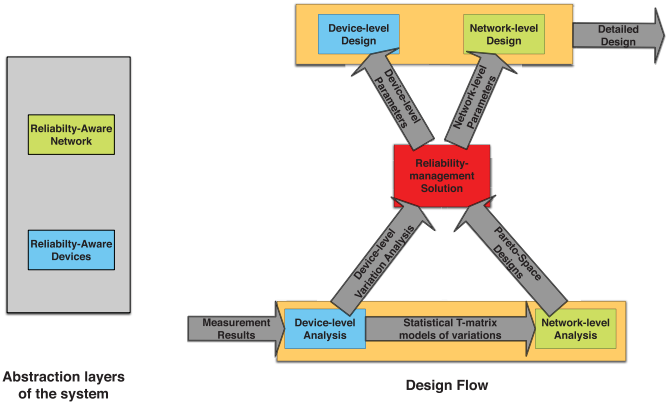

Fig. 3. Reliability-aware design flow based on abstracting our system into two levels.

III. Reliability-Aware Flow

To address the challenge of silicon photonic interconnect, a reliability-aware design flow is proposed. As shown in Fig. 3, there are three steps in this flow: analysis, management, and design. During analysis, a fabrication-calibrated device model quantifies the impact of variations on device reliability. Based on this model, a Pareto-space of device and network designs are explored. The impact of variations on system reliability is minimized through choosing optimal design points. Due to the absence of variation-free design, a light-weight reliability management solution is supplied based on the optimal choice of device and network. Detailed design of devices and network is then conducted to complete the flow.

The flow is based on an abstraction of the interconnect. The interconnect is separated into two tightly coupled levels: device and network, as shown in Fig. 3.

1) The device level consists of individual devices and point to point optical links. At this level, variations and their impact on the device response are quantified statistically.
2) The network level consists of network architecture. The architecture is constrained by performance, power, and reliability. Device models are used quantitatively to estimate system variations. Based on these models, architecture-level simulation is used to explore the Pareto-space of design.

The device- and network-levels are tightly coupled to address reliability challenges. Device-level design requires network-level information. For instance, WDM may employ a wideband directional coupler or narrow-band microrings.

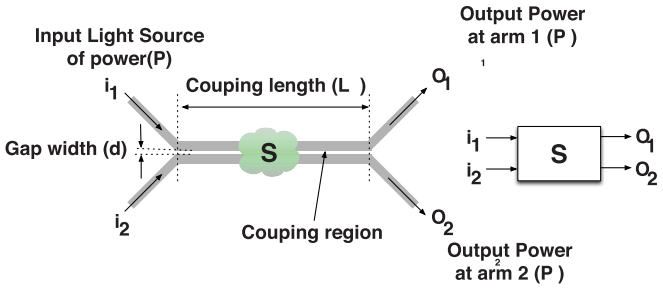

Fig. 4. T-matrix of directional coupler.

The performance requirements and area constraints determine the device design.

We use yield as a measure. Yield is defined as the number of chips that conform to the design specifications and are reliable (exhibit bit-error-rate of $<10^{-12}$). Spectral overlap of individual wavelength channels, for example, is an important component of BER. Mathematically, yield, Y, as limited by wavelength drift can be expressed as

$$Y = \int_{\mu-\lambda_t}^{\mu+\lambda_t} \mathrm{e}^{-\frac{1}{2}\left(\frac{\lambda-\mu}{\sigma}\right)^2} \mathrm{d}\lambda \tag{4}$$

where μ is the mean resonant wavelength, σ is standard deviation of resonant wavelength variations, λ is wavelength, λ_t is half the tuning range.

The yield numbers presented here are based on our own measurements of our own fabrication results. In the following, we study power and performance design metrics in detail.

A. Device-Level Reliability-Aware Design Flow

In this section, we discuss the device-level analysis and design steps in the flow. We present a fabrication calibrated reliability-aware model for analysis and design. During analysis, the model is to determine the statistical behavior of the model parameters. During design, the model is used to evaluate device response when the design parameters vary. We present a case study on the design of a 16 WDM microring-based communication link with predetermined yield.

1) Reliability-Aware Device Modeling and Analysis: In this section, we develop a reliability model. We use transfer-matrix (T-matrix) as a phenomenological tool. We illustrate the approach for two filters, the directional coupler and the microring.

A T-matrix model describes the power transmission spectrum of a multiport device. T-matrix model for a directional coupler is illustrated in Fig. 4. A directional coupler consists of two symmetric waveguides with a gap, d_g between them. When light propagates in one arm, part of the input power is coupled to the second waveguide. The amount of power coupled from one waveguide to another depends on the gap width (d_g) and the coupling length (L_c) at the specific wavelength (λ). The transmission curve in terms of resonant wavelength is a cosine function. A two-input two-output

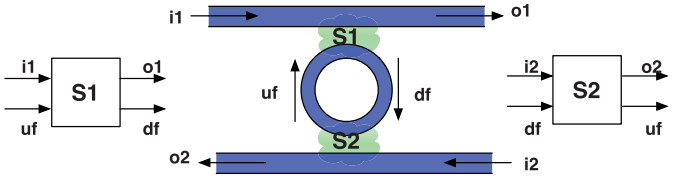

Fig. 5. T-matrix models of microring resonator.

directional coupler exhibits wavelength spectrum [14]

$$\begin{pmatrix} o_1 \\ o_2 \end{pmatrix} = \begin{pmatrix} r & -it \\ -it^* & r^* \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} \tag{5}$$

where i_1 and i_2 are the input ports, o_1 and o_2 are the output ports, t is the coupling coefficient, whereas r is the energy that stays within the same waveguide. r^* and t^* are the complex conjugates of r and t

$$r = \cos\left(L_c\sqrt{\kappa^2+\delta^2}\right) + i\frac{\delta}{\sqrt{\kappa^2+\delta^2}}\sin\left(L_c\sqrt{\kappa^2+\delta^2}\right) \tag{6}$$

$$t = \frac{\kappa}{\sqrt{\kappa^2+\delta^2}}\sin\left(L_c\sqrt{\kappa^2+\delta^2}\right). \tag{7}$$

In (6) and (7), L_c is the length of the two parallel coupling waveguides, κ is the coupling coefficient that determines the coupling efficiency. δ is defined as $\Delta\beta(\lambda)/2$ where $\Delta\beta$ is the propagation constant difference between the first and the second modes that are in the coupling region S.

The model parameters include both design parameters and fabrication-specific parameters. For instance, L_c is a design parameter controlled by the designer, on the other hand, δ is fabrication-specific parameter (and also depends on the gap parameter determined by designer).

Fabrication-specific parameters are lumped with other parameters in conventional methods such as computation-intensive electromagnetic simulations. For instance, the change in temperature, process variations in gap, and process variations in width and thickness of the waveguide are all contained implicitly in the analytical variable δ.

Design parameters provide powerful tools to explore the design space of the device with constraints. For instance, by varying the coupling length L_c, one can get the wavelength spectrum of different designs.

The T-matrix model for microring can also be used for variability analysis. A microring in Fig. 5 can be modeled as a concatenation of T-matrices of the form [14]

$$\begin{pmatrix} o_1 \\ d_f \end{pmatrix} = \begin{pmatrix} r & it \\ it & r \end{pmatrix} \begin{pmatrix} i_1 \\ e^{i\phi/2}\mathrm{e}^{-\alpha l/2}u_f \end{pmatrix} \tag{8}$$

$$\begin{pmatrix} u_f \\ o_2 \end{pmatrix} = \begin{pmatrix} r & it \\ it & r \end{pmatrix} \begin{pmatrix} \mathrm{e}^{i\phi/2}\mathrm{e}^{-\alpha l/2}d_f \\ i_2 \end{pmatrix} \tag{9}$$

where i_1 and i_2 are the input ports, o_1 and o_2 are the output ports, d_f is the downward stream in the microring, and u_f is the upward stream in the microring as shown in Fig. 5, α is the loss per unit length and l is single round trip length, ϕ is the phase progression from a round trip, which can be expressed as by $\phi = 2\pi n_{\mathrm{eff}} l/\lambda$, where n_{eff} is the effective refractive index and varies with applied voltage, device geometry, and the ambient temperature, and λ is the wavelength in free space.

In (8) and (9), we assume the coupling regions (S_1 and S_2 in Fig. 5) are symmetric and (almost) lossless. t is the field coupling coefficient from the waveguides into the microring, and also from the microring to the waveguides. r represents the amplitude of the remaining field. Hence, we have $r^2+t^2=1$.

Although we focus on yield as our figure of merit, other figures of merit are also possible such as area, transmission efficiency, bandwidth as [15] shows. There are also other analytical approaches available. We adopt the T-matrix model to predict yield for three reasons.

1) T-matrix contains the fabrication parameters explicitly. The model parameters can be obtained from fabricated devices by measurement calibration. That is, we can extract α, n_{eff} and coupling coefficients from the measurements. By fitting the responses of multiple fabricated devices of the same nominal design at different temperatures, we can extract the effects of fabrication inaccuracy and thermal variations directly. The accuracy of extraction is much higher than what can be obtained from time consuming electro-magnetic simulation. From our analysis, we find that n_{eff} contains valuable information about waveguide thickness and width variations and the thermo-optic coefficient responsible for thermal sensitivity.
2) Statistical analysis of calibrated parameters provides the foundation of the reliability-aware device design. As depicted in Fig. 6, we can calculate statistics and determine which statistics have Gaussian distribution, by comparing the model with measurement. We can then find means and standard deviations. These statistical models form the bases of our analysis as will be illustrated in the design of a 16 channel microring-base WDM filter. For example, knowing that n_{eff} in microrings have a mean of μ_n and standard deviation σ_n, one can analyze the variations for microrings of different radii and gaps to get a mean resonant wavelength of μ_r and a standard deviation of σ_r for different designs with high accuracy and low computational cost. A microring with a standard variation of 2.59 nm on a waveguide line-width [13] would result in a 2.41 nm shift in resonant peak; hence, a yield of 46% in a 64 channel system.
3) Device-level information can be translated to network-level through the calibrated T-matrix models. Modeling a set of connected devices that represent a network on-chip can be done by multiplying the T-matrices models. This network-level analysis can provide the statistics of the signal-to-noise ratio and bit-error-rate reliability metrics. Optical and electric power and analytical performance metrics can be computed for a whole network. Considering an optical link building with the microring in the last example. The 46% yield drops to 7% for the link. The 7% is too low to build a system. Hence, a design space exploration of the network is possible through low-computational overhead matrices multiplications [15].

2) Reliability-Aware Device Design: In this section, we discuss reliability-aware device design techniques taking into account variations and yield. We leverage the analysis results

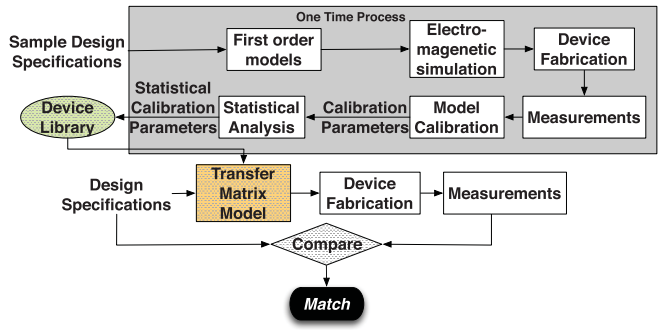

Fig. 6. Proposed design process for Silicon Photonic devices.

at both the device- and architecture-levels for detailed specifications of the device characteristics such as full-width-half-maximum, free spectral range, and resonant wavelengths. We propose a new design process that is superior to current slow and inaccurate process. We demonstrate this design process through an example of a 16 WDM channel microring based multiplexer.

Traditional techniques for design of silicon photonics devices are slow, iterative, expensive, and do not explicitly account for variations. The traditional design process starts with a wavelength spectrum of a device specification, which are then translated to design parameters using first order models. Next, simulations are conducted in an iterative approach to fine tune the design parameters. If process and thermal variations are to be taken into considerations more simulations for each variation point is required. This iterative procedure is long where a single iteration can take up to 10 h depending on how complex the structure under study is. After that, the design parameters are available and the device is laid out for fabrication. The fabrication process is an expensive and long step. Despite this effort, the output device when measured does not match the device specifications. This returns to several reasons such as process variations and inaccuracies in simulations. Current electromagnetic simulations present an idealized framework for simulation without accounting for real fabrication effects such as surface roughness and impurities introduced in the fabrication process. Hence, the fabricated device does not meet our device wavelength spectrum. This has motivated us to pursue alternative design processes as we will present here.

We propose a novel reliability-aware design process. The new design process is a two step process as shown in Fig. 6. In the first step, our goal is to calibrate the fabrication-specific parameters of the T-matrix model discussed in Section III-A.1, statistically analyze the model parameters to compute mean and standard deviation, and build a library of devices.

The second step involves the design of the target device with a predefined wavelength response spectrum to conform to the design specifications the analysis phase and reliability management solution recommends. For instance, a directional coupler design would involve the specification of wavelength of peak transmission and the period of the peaks. In the model, we have the fabrication-specific parameters and design parameters. We leverage the device library that includes the

fabrication-specific parameters for our design and tune the design parameters to get the required response. This design process is more accurate than long electromagnetic simulations since it accounts for process-specific effects.

a) Case study: design of 16-channel WDM multiplexer: The present case is the design of a 16-channel WDM multiplexer. The experimentally fixed model parameters are obtained from a fabricated 4-channel WDM multiplexer. The target 16-channel multiplexer has tighter inter-channel spacing than the 4-channel. The fabricated 4-channel WDM design is composed of microrings of different radii. The microring radii are 4.98, 5, 5.02, and 5.04 μm. Measurement results indicate that the channel spacing has 2.45 nm mean value and a standard deviation of 0.836 nm. This design was fabricated and measured 12 different dies with two devices per die, where all dies were on the same wafer. Out of the 24 structures measured, in only one instance did adjacent WDM channels overlap. The yield then is 96% and is expected to be more than 99% yield based on our standard deviation of channel spacing variations for large sample count.

The design of the 16-channel WDM multiplexer follows the following steps.

1) *Design goal:* We target a 16 WDM channel multiplexer and a yield of 68.27% (one standard deviation spacing). The channels should be designed to fit into the minimal bandwidth required for the yield and channel count.
2) *Design parameters:* The T-matrix model design parameters are effective index n_{eff} and radius R. The designer has less control over the effective index than radius. Waveguide dimensions are designed for minimal loss (and least sensitivity to process variations). The target channel spacing is required to be 0.84 nm in order to realize our 68.27% yield. We need to choose radii R accordingly.
3) *Design process:* The T-matrix models dictates that the relation among resonant wavelength λ, effective index n_{eff}, and radius R is $m\lambda = 2\pi n_{\text{eff}} R$.

Expanding for small wavelength deviations between channels, one finds

$$m\Delta\lambda = 2\pi\ (R\Delta n_{\text{eff}} + n_{\text{eff}}\Delta R) \tag{10}$$

where $\Delta\lambda$ is the channel spacing, ΔR is difference in radius for different microrings in the structure, and Δn_{eff} quantifies the process variations in waveguide thickness and width.

Plugging in for Δn_{eff} and n_{eff} from our measurement results and $\Delta\lambda$ for our target design, we find that the radius change between subsequent microrings in the structure ΔR (which is 14 nm on average) leads to an average spacing between channels of 0.835 nm. The T-matrix model leads to a new 16-channel WDM filter design of predetermined yield in advance with little computational overhead in comparison with electromagnetic simulation.

B. Reliability-Aware Device Management

Existing reliability management techniques are primarily functioning at device-level. They are neither cost-effective nor robust. We have conducted a survey of device-level reliability management approaches. We will presently discuss these techniques from a system perspective and evaluate them from the point of view of on-chip optical network design. We compare of different techniques for variation-resilient design in Table I in terms of their capabilities and overhead. We have categorized these techniques based on their targeted variations: athermal techniques and process-variation-immune techniques.

1) Athermal Silicon Photonic Devices: Herein, we present the various device-level techniques to mitigate thermal variations. We focus on techniques that apply to microring devices. The three main methods are: 1) tuning the characteristics of the microring; 2) using polymers to counteract the effect of silicon's high thermo-optic coefficient; and 3) miscellaneous.

First, we discuss tuning based approaches.

1) *Voltage tuning:* The resonant wavelength of a microring can be tuned by carrier injection/depletion that alters the carrier concentration in the silicon waveguide and blue/red shifts in resonant wavelength [16]. The tuning range is limited to a few nanometers because of exponential increase in free carrier absorption with injection [6], [7]. Manipatruni et al. [16] demonstrated athermal microring operation for a temperature range of 15° that corresponds to 1.65 nm.
2) *Bandwidth tuning:* By restructuring the coupling region and using the thermo-optic effect of silicon, one can tune the bandwidth of the microring from 0.01 nm to a few nanometers. Chen *et al.* [17] demonstrated the technique through fabrication and measurements where they achieved a tuning range of 0.1-0.7 nm. By increasing the bandwidth of the microring, shifts in resonant wavelength can be tolerated. However, the bandwidth is still limited and tuning range to less than 10 °C.
3) *Thermal tuning:* In this approach, a microheater heats the microring. The heating red-shifts the resonant wavelength. Several groups have demonstrated thermal tuning across a wide range of wavelength [18]. The system application lies in counteracting a red shift in resonant wavelength at one side of a link by heating the microring on the other side. This approach has a negative impact on the system from two perspectives: first, it greatly increases the total power consumption of the system [6], [7] and second, it increases the temperature of the system which affects overall reliability in many-core systems [7].

The second class of approaches leverage thermal properties of polymer coatings. Polymer coating of devices involves an extra post-processing step wherein the chip is spin coated with polymer that is subsequently baked. This kind of processing is feasible and low cost. We have demonstrated such post processing in our lab for various types of polymers [25]. The cost overhead is minimal. Polymer coating requires no control overhead at run-time. The idea is to guide part of the light in a polymer that has the opposite sign of thermal coefficient to silicon. The overall propagation constant that is the average of that in the silicon and that in the polymer can then be made athermal if a polymer with the correct thermal coefficient is compatible with the process. This can be achieved through use of: 1) polymer cladding with negative thermo-optic coefficient;

TABLE I
ATHERMAL DEVICES CHARACTERISTICS

Technique	Extra Optical Loss	Electrical Power	Switching Speed	Post-fabrication Step	Reliability	Additional Area	Sensitivity	Temperature Range	Wavelength Range
Voltage Tuning [16]	2-5 dB	4.52 mW/bit	1 Gbps	Doping	High	-	-	15 °C	C+L band
Bandwidth Tuning [17]	0 dB	26.4 mW/bit	few Mbps	Add Heater	Low	-	-	10 °C	C+L band
Thermal Tuning [18]	0 dB	0.2 mW/bit	>6 Kbps	Add Heater	Low	-	-	0-104.5 °C	C+L band
Polymer Cladding [9]	0 dB	0 mW	Gbps	Polymer coating	High	-	0.5 pm/°C	25-125 °C	1450-1550 nm
Slotted Device [19]	5-8 dB	0 mW	Gbps	Polymer coating	High	-	52 pm/°C	20-65 °C	1524-1576 nm
All-Polymer waveguide [20]	0 dB	0 mW	unverified	CMOS-Incompatible	High	several mm^2	-0.9 pm/°C	20-65 °C	1550 nm
Couple to MZI [21]	3 dB	0 mW	Gbps	-	High	$230\times796\ \mu m^2$	-	20-100 °C	1565.5 nm
Stress Control [22]	unknown	0 mW	Gbps	Several steps	Low	-	1.8 pm/°C	33-90 °C	1550 nm
Micro-fluid Tuning [23]	unknown	0 mW	slow	Bond Micro-fluid Chip	Low	-	-	9 °C	1552.11 nm
Evanescent Field Perturbation [24]	unknown	0 mW	slow	-	Mechanical	Probe area	-	>200 °C	1560-1595 nm

2) low thermal coefficient polymer filled slotted structures—a problem being that slotted structures exhibit higher loss than ridge structures; or 3) all polymer waveguides (no silicon involved) with negligible thermal optic coefficient that do not exist and would be weakly guiding (much larger than silicon photonic structures).

Our third class of athermal microrings include a miscellaneous set of approaches; however, they are less promising than the earlier proposed techniques, and we include them for completeness in Table I.

In conclusion, different techniques come with different overheads or limitations. As shown in Table I, the athermal polymer techniques are the most promising techniques to eliminate thermal variations. Moreover, they have little overhead. However, using polymer for thermal variations does not allow us to use polymer for controlling process variations as we discuss in the next section.

2) Process-Variation-Immune Silicon Photonic Devices: Techniques to overcome process variations overlap with athermal device techniques. For example, the same tuning techniques used in athermal microring design Table I could be used for counteracting process variations. The techniques include thermal tuning, bandwidth tuning, and voltage tuning. An important distinction is that the tuning for process variations need be done but once. There is no need for run-time management.

Trimming can be used for process variations. In trimming, the microring or silicon photonic device is coated with a ultraviolet (UV) sensitive polymer overlay. The device spectrum is measured. The polymer is exposed either during measurement to a dose of UV radiation (or in other cases photo-bleaching) until the target properties are obtained [26]. The challenge here is the real time aspect. Each microring exhibits a different range of process variations and requires a different dose of UV. Individual addressing may be complicated in a complex circuit. Use of a different mask for each microring is, in general, labor and cost intensive. One could limit the number of exposures by assuming that the magnitude of the process variations vary slowly across a wafer. Many measurements and many exposures would still be necessary.

Based on our yield model in (4), we now compare device-level techniques: Thermal tuning can totally eliminate the impact of variations with a high power and switching speed cost, giving a yield of 100%. Voltage tuning can eliminate the impact of process variations with a yield of 82% and thermal variations with a yield of 68%, in both cases with a lower power cost. However, voltage tuning cannot simultaneously be used to reduce both types of variations. Athermal polymer coatings can eliminate the run-time overhead with a low cost but cannot counteract process variations giving a yield of 82%. Finally, trimming can reduce process variations but not thermal variations giving a yield of 68%. No single solution can solve the reliability problem without significant overhead. Hence, a system-level view is necessary. Next, we demonstrate the tradeoffs involved of different device-level reliability management techniques through a case study.

3) Case Study: Reliability-Aware Design and Management of a Single Channel Optical Link: In this paper, we design a single channel WDM optical link that can operate reliably despite process variations. We leverage our analysis of process and thermal variations for different WDM-structures including microrings, racetracks, and directional couplers to make device design decisions and reliability management decisions.

In our analysis we account for power and yield. Our process variations analysis is for specific design parameters of the devices considered. We focus on the following devices in our analysis.

1) *Microring:* The microring has a radius of 4.98 μm and a gap of 200 nm. The design was fabricated and measured on 12 dies, replicated two times in the same die, on the same wafer. The wafer was fabricated in LETI [12].
2) *Racetrack:* The racetrack has a bending radius of 3 μm, a coupling length of 7 μm, and a gap of 130 nm (the minimal feature size). The design was fabricated and measured on 18 dies on the same wafer. The wafer was fabricated in IMEC [12].
3) *Directional coupler:* The directional coupler design we use has 130 nm gap (the minimal feature size) and length of 1063 μm. The design was fabricated and measured on 18 different dies on the same wafer. The wafer was fabricated in IMEC [12].

To design a reliable optical link with minimal power overhead requires one to account for the following factors.

1) *Variation in resonant wavelengths:* Reliability analysis provides the standard deviation in resonant wavelength variations due to process and thermal variations. The required tuning power is determined by the magnitude of these variations. For the three types of devices, a summary of our variations analysis is provided in Table II.

TABLE II
RESONANT WAVELENGTH, PROCESS, AND THERMAL VARIATIONS ANALYSIS OF THE DEVICES [5]

Device	FWHM (nm)	FSR (nm)	Process Variations μ	Process Variations σ	Thermal Variations nm/°C
Micro-ring	0.65 nm	16.5 nm	1534.4 nm	1.2 nm	0.11 nm/°C
Racetrack	0.85 nm	15.7 nm	1533.3 nm	2.16 nm	0.09 nm/°C
Directional Coupler	3.56 nm	6.7 nm	1550 nm	1.12 nm	0.12 nm/°C

2) *Channel full-width-half-maximum (FWHM) and free spectral range (FSR):* Both FWHM and FSR impact the sensitivity of device operation to variations. The larger the FWHM of the resonant peak, the more tolerant the channel is to variations. The shorter the FSR, the less tuning range. Among the three types of devices, the directional coupler has widest FWHM and smallest FSR implying lowest tuning range, followed by racetracks, and finally microrings as shown in Table II.

Next, we compare the different device-level reliability management techniques for an optical link given the design goal (minimizing tuning power for a given yield) and design parameters (tuning technique and device choice). Different tuning techniques provided in Table I have different overheads. We focus on voltage tuning and thermal tuning as in the first portion of Table I since they can compensate both process and thermal variations. The comparison is based on the electrical tuning power computed for each device and tuning technique for both average and worst cases. The worst case puts an upper limit on power dissipation, while the average case gives a statistical value for power dissipation that is useful for multichannel network with multiple links. Our models from Section III-A provide the variation ranges for effective index due to process and thermal variations and consequently a simple analysis (from phase 1 of our design flow) will give us the resonant wavelength of the different devices. According to our calculations, the minimal overlap is 0.5 nm for a light source of 780 μW/nm and an optical path loss of 10 dB. Hence, the total worst and average statistical case power is given by

$$P_{\text{total}} = P_{\text{tuning}} \times \frac{(\text{FSR} - 2 \times \text{FWHM} + 1)}{2} \tag{11}$$

$$P_{\text{total}} = \frac{2P_{\text{tuning}}}{\sigma\sqrt{2\pi}} \int_{\mu}^{\mu+\sigma} (\lambda - \mu)\, \mathrm{e}^{-\frac{1}{2}\left(\frac{\lambda-\mu}{\sigma}\right)^2} \mathrm{d}\lambda \tag{12}$$

where P_{total} is total tuning power for respective tuning technique, P_{tuning} is tuning power per nanometer, FSR is free spectral range, FWHM is full-width half maximum, μ is mean of resonant wavelength, σ is standard deviation of process variations, and λ is resonant wavelength.

Equation (11) assumes worst case tuning, that is, where the resonant peak of the transmitter lies midway in the free spectral range of the receiver. Meanwhile, (12) assumes a Gaussian distribution for different devices, and the average tuning power is calculated through integrating the power per nanometer across resonant shift in the distribution.

TABLE III
THERMAL TUNING POWER OF DIFFERENT OPTICAL LINKS

Device	Thermal tuning					
	Average case			Worst case		
	Tuning Range	Power	Yield	Tuning Range	Power	Yield
Micro-ring	1.06 nm	0.031 mW	68.27%	8.1 nm	1.62 mW	99.9%
Racetrack	1.81 nm	0.051 mW	68.27%	7.5 nm	1.5 mW	99.9%
Directional Coupler	0 nm	0 mW	68.27%	0.29 nm	0.058 mW	99.9%

TABLE IV
VOLTAGE TUNING POWER OF DIFFERENT OPTICAL LINKS

Device	Voltage tuning					
	Average case			Worst case		
	Tuning Range	Power	Yield	Tuning Range	Power	Yield
Micro-ring	1.06 nm	0.0236 mW	68.27%	8.1 nm	-	-
Racetrack	1.81 nm	0.039 mW	55.5%	7.5 nm	-	-
Directional Coupler	0 nm	0 mW	68.27%	0.29 nm	0.0446 mW	99.9%

The tuning range, power, and yield results for (4), (11), and (12) for the three different device design under study and the two tuning techniques are provided in Tables III and IV. As one can see, voltage tuning can tune the process variations within one standard deviation for the different devices but for a range >1.65 nm, the optical loss is too high as is the case also for microrings and racetracks; hence, voltage tuning cannot be used leading to lower yield levels. On the other hand, thermal heating can tune any range of variations leading to higher yield but at slightly higher power levels (and at expense of post-processing the wafer for creating undercut structures and for low switching frequencies). In terms of device choice, directional couplers dominate, due to their large FWHM and small FSR, making them immune to process variations within one standard deviations and very low tuning ranges for worst case tuning range. The yield for directional couplers is optimal as for both average and worst case tuning ranges.

In this paper, we have designed a variation-aware single channel optical link using our analysis of variations of fabricated devices and our models for process variations. Since we are considering a single link, power is the main design criteria but at higher bandwidth the tradeoff between power and bandwidth needs to be considered.

C. Network-Level Reliability-Aware Design Flow

The second abstraction level in our design hierarchy is the network-level. This level is essential in analyzing the performance of the network under realistic variations and loads, defining a Pareto-space of design points, and detailed design of the network in terms of topology, optical power, and bandwidth to meet the power and performance requirements of the system. In this section, we present the result of our analysis phase simulation in terms of design guidelines.

The design parameters that the architect has control over are: optical path design, power budget, and bandwidth. These parameters need to be fixed in light of expected variations. For example, the loss along each optical path needs to be selected such that the network will operate reliably under a

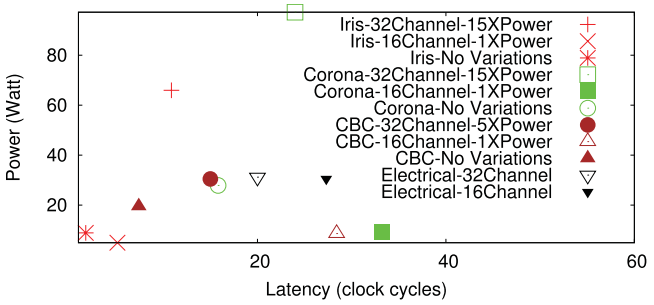

Fig. 7. Design space exploration in the architecture domain.

variety of workloads. As we will see, worst case design leads to unnecessary overhead. Worst cases are best treated with active control.

The design parameters need to be studied in light of the design goals. Due to the strong correlation between the different design goals: 1) performance; 2) power; and 3) reliability, designing in this 3-D space becomes more complicated and a simulation framework is necessary to show the architect the Pareto space of design points. Simulation frameworks based on reliability models provide analysis of performance in terms of latency and bandwidth, power in terms of optical power and electrical power, reliability in terms of bit-error-rate.

The goal of this design space exploration step is twofold: first, to assess our architecture design options; second, to quantitatively understand what freedom the power budget and performance constraints allow. Next, we demonstrate how to perform an analysis step and produce the Pareto-space of design points as illustrated in Fig. 7.

Our simulation analyzes the impact of variations on the network and defines the design space. We simulate five different networks: 1) electrical mesh network; 2) iris [2]; 3) corona [4]; 4) macro-chip network [27]; and (5) channel-borrowing crossbar (CBC) [28]. A benchmark ("lu") in SPLASH2 [29] is used as an example. We simulate a 64-node network which has a spatial thermal variation of 15° and process variations with a standard deviation of 1.2 nm in resonant wavelengths. Such assumptions are based on the fabricated device measurement analysis conducted in our lab. We leverage our simulation framework that models performance, power, and, reliability, where error detection and retransmission is assumed to accommodate communication errors. Our simulation framework is a network-on-chip simulator that accurately models the network assuming a baseline network configuration described in Table V. The simulation framework is flexible enough to allow the control of the design parameters on the network in terms of performance and power. Three parameters were varied during simulations: 1) Optical path design through choosing different networks discussed earlier; 2) The bandwidth per link (number of bits that can be transmitted concurrently) where we considered 64, 32, 16, and 4 channels per link for the optical network (through different resonators including microrings, racetracks, and directional couplers), and 32, 16, and 4 bits links in the electrical network; and 3) Input optical power for silicon photonic networks by increasing the power by a factor of 2×, 3×, 5×, 10×, and 15×. Fig. 7 shows the results of the networks within an acceptable power and performance design space (other results beyond this range are omitted for clarity). In Fig. 7, the power and latency (our design goals) are the axes of the design space, while the optical path design, optical power, number of channels (our design parameters) are represented through different network designs with different power and bandwidth and marked by different markers in the graph. From this paper, we can verify our design guidelines as follows.

1) Optical Path Design (Topology and Flow Control): Topology and flow control are crucial to the tolerance of the network to variations. Topologies to date are one of five categories: 1) Single-Write-Multiple-Read (SWMR) networks that leverage a serpentine link that broadcasts data [30]; 2) Multiple-Write-Single-Read (MWSR) networks, also called crossbars that leverage a shared serpentine link to deliver data to a single receiver [4]; 3) Repeated link networks which exhibit short optical links where the optical signal is regenerated through electrical routers [31]; 4) Antenna-based broadcast such as Iris [2]; and 5) Optical switch-based designs such as the mesh network in [3] and Iris [2]. First, repeated link networks are favored over SWMR and MWSR networks due to the short optical paths with lower optical loss, where high signal power levels tolerate higher noise and variation levels. Optical switches are a major source of crosstalk in optical networks [6]. Switches provide flexibility in network design, but the number of switches along an optical path should be minimized through inspired design. Broadcast-based and SWMR networks mandates that all nodes in the system are matched since they are all participating in the communication. This adds additional constraints from a reliability perspective. Our simulations confirm the necessity of these guidelines through the power and latency results provided. The Macro-chip network survived the process variations at high channel count (32 channels) due to the very short optical path. However, the latency was too high to include in the design space due to the small bandwidth allocated per core-to-core communication link. Meanwhile, broadcast-based networks such as Corona (Corona-32Channel-15 × Power and Corona-16Channel-1 × Power in Fig. 7) suffer from sensitivity to variations, Here, any mismatch of one of the communicating nodes leads to retransmission resulting in the observed high latency of these networks. Finally, unicast networks such as channel-borrow crossbar (CBC-32Channel-5 × Power and CBC-16Channel-1 × Power in Fig. 7) are less sensitive to variations and can operate at lower optical power levels but are more sensitive to bandwidth variations.

2) Bandwidth Design: The number of WDM channels is another factor in the design decision. The greater the number of WDM channels used, the smaller the bandwidth per channel. Narrow channels imply less tolerance to variations. Assume a channel of bandwidth 3 nm for a 32 channel network. Such a network can tolerate up to a 3 nm passband shift that is equivalent to a 27° thermal variations. On the other hand for a 128 channel network can tolerate only 0.75 nm that is equivalent to a 7° thermal variations. However, the 128 channel system has higher bandwidth and better performance under ideal conditions. Our simulation study yields further details as follow: At high bandwidth (64 channels) all

optical networks fail and at low channel count (four channels) the variations have negligible influence but the bandwidth is so small that the latency deteriorates beyond our acceptable design space. At nominal power levels, at most 16 channels (like Iris-16Channel-1 × Power, Corona-16Channel-1 × Power, and CBC-16Channel-1 × Power in Fig. 7) can be used to accommodate the variations in the system. However, the performance suffers from temporal variations (some packets cannot be received) and low bandwidth.

3) Error Detection and Correction: Redundancy in data transmission that result from additional error detection and correction bits can greatly improve reliability. Nitta *et al.* [32] propose using error detection codes and forward error correction codes to detect and correct different types of errors in the network including inter-channel crosstalk, partial and total passband mismatch. The number of data channels used for error detection and correction is a tradeoff between reliability and performance.

4) Power Budget Design: The input laser power of the network plays a major role in reliability. Increasing the input laser power increases the signal-level, and consequently improves signal-to-noise-ratio and bit-error-rate. Partial mismatch between sender and receiver(s) can be mitigated by a higher power level. The drawback is the increase in the power budget, a major design criteria, and the introduction of thermal reliability problems to the CMOS many-core layer. A tradeoff between power and reliability arises. Careful design is necessary to balance the two demands [8]. Quantitatively, our simulation results verify this as follows: At 32 channels, a 15× optical power increase for broadcast-based networks (like Corona and Iris which correspond to Iris-32Channel-15 × Power and Corona–32Channel-15 × Power in Fig. 7) and 5× optical power increase for other networks (like Channel-Borrow crossbar which correspond to CBC-32Channel-5 × Power in Fig. 7) is needed to reduce the latency to our acceptable design space in Fig. 7.

From the above paper, we conclude that Iris dominates in terms of power and performance but at a higher cost of extra 3-D stacked layers, while Corona and channel-borrow crossbar offer a Pareto space of possible designs. Electrical networks cannot compete with optical counterparts. However, the study also shows that variations are problematic for nanophotonic networks on-chip. Iris represents best power-performance but has a 3× higher latency when compared to the ideal, variation-free case. Silicon photonics promises low latency at low power levels as earlier studies in Iris [2] for 64 cores and ATAC [30] for a thousand core indicate. New techniques to mitigate the impact of variations are necessary to realize these improvements in practice.

D. Reliability-Aware Network Management

Similar to device-level reliability management, different network-level reliability management techniques have been developed. In this section we review different network-level reliability management techniques while assessing their capabilities and limitations. The survey shows that these techniques do not solve the reliability problem despite their overhead.

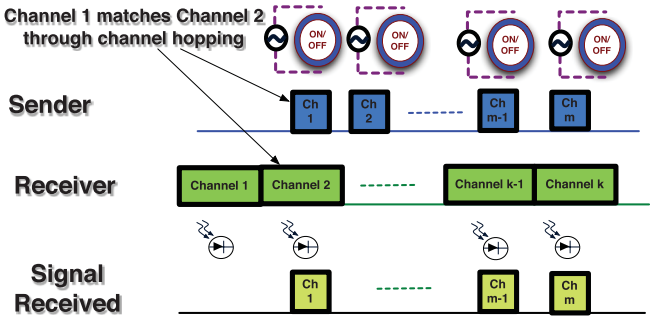

Fig. 8. Channel hopping technique for coarse grain passband matching.

Careful design and trade-offs combined with device-level reliability management techniques are necessary to design a reliability management solution. Next, we discuss the list of network-level reliability management techniques:

1) Channel Hopping: At the coarse grain level we can leverage channel hopping to address mismatch between sender and receiver: Due to variations, a microring at the sender and one at the receiver designed to have matching passbands will mismatch. However, the sender microring for channel k will overlap with the receiver microring of channel m. Moreover, due to locality of process variations, channel $k+1$ will overlap with channel $m+1$. Consequently, we redefine the channels in terms of overlapping fabricated passband not the overlapping design passband, as shown in Fig. 8. Through channel hopping, we managed to communicate a signal between the sender and receiver despite the mismatch due variations. The area and power overhead of this approach comes from the extra microrings at receiver to guarantee that each channel at the sender has one matching microring at the receiver.

2) DVFS: Leveraging the thermal profile of the chips to counteract process variations in the network has been proposed by Li *et al.* [7]. The goal is to control the voltage and frequency of the different cores to generate a spatial thermal profile that counteracts the process variations of the microrings serving the corresponding core. Two limitations come to our attention: 1) The thermal profile is coarse grain and not uniform across the whole core; hence, fine grain tuning is necessary and 2) The thermal profile is not temporally uniform and can fluctuate by 5° [33].

3) Workload Migration: Similar to DVFS, workload migration can be used for thermal management; however, this approach cannot be used to generate a thermal profile. The studies by Li *et al.* [7] and by Oh *et al.* [34] show that workload migration is a strategy to stabilize and not to control temperature. Hence, it is augmented with DVFS to stabilize the thermal profile, reduce the DVFS overhead, and thus improve the throughput of the whole system.

IV. Reliability-Aware Management Solution

In this section, we show how to combine the reliability management flow to make design decisions for a reliable, high-performance, and low-power system. The proposed solution is for our specific application (architecture and workload). Other solutions are possible for different systems through the same design flow proposed herein. The purpose of the proposed solution is to show how effective our design flow is by

demonstrating the improvement in the reliability management solution in various aspects. In this reliability management solution, we make six crucial design decisions as follows.

1) *Number of channels:* Based on our architecture-level study, we have decided on 64-WDM channels through using racetracks and microrings.
2) *Polymer cladding:* Polymer cladding is chosen to address thermal variations induced reliability challenge due to its low run-time overhead compared to other techniques.
3) *Channel hopping:* To counteract process variations, we address this problem at a coarse grain level through channel hopping. For a 64 channel system and 1.2 nm within-die process variations, one extra microring at each side are enough which is equivalent to 3% area overhead at receiver.
4) *Voltage tuning:* For fine-grain passband tuning (within one channel bandwidth) we leverage voltage tuning to align each sender with only one receiver.

We propose to have coarse-grain and fine-grain tuning be done as a post-fabrication step. This can be done as follows: we sweep a narrow-band laser (the passband is equal to the sender microring passband) over the entire wavelength range to give us the resonant wavelength of different microrings in the sender by detecting which receiver has received a signal and establish a communication channel. Consequently, coarse grain tuning has been achieved. If inter-channel crosstalk occurs where two detectors receive a signal from same sender voltage tuning is leveraged in this case to tune the sender to one of the receivers. The configuration is done once in a post-fabrication step with minimal circuitry overhead since the communication channels are established statically in a one-time step. Voltage tuning configuration is also configured once through external signal driving the control circuitry. No dynamic configuration overhead incurs and no additional power or circuit complexity is needed for this solution.

Network Design: Our system-level solution is independent of the underlying network architecture, which provides flexibility to the network architect to make design decisions. According to our study and Pareto-space from our analysis in Section III-C, we have focused on broadcast-based solutions, more specifically Iris [2] due to its low latency and low power characteristics compared to other solutions.

Variation-Aware Design: In this solution, we opt for designing to have a yield of 96% which corresponds to two standard deviations of our Gaussian distribution, not worst case design. Worst case variations can be very costly in terms of power and area as we show in Section V.

V. Simulation Results

In this section, we evaluate our proposed system-level reliability management solution in terms of power and performance. We compare our system-level solution to other system-level solutions such as the solution in [7] that leverages Dynamic Voltage and Frequency Scaling (DVFS), workload migration, and voltage tuning, the solution in [6] that leverages channel hopping, voltage tuning, and run-time routing, the solution in [10] that leverages channel hopping, voltage tuning and thermal tuning, the solution in [11] that leverages redundant channels and thermal tuning, and a solution similar to our solution but with accounting for worst case variations instead of statistical variations.

TABLE V
Configuration of Our Simulation Framework [2]

Nanophotonic network in Iris		Memory Hierarchy and Processors	
Throughput	4Gb/s	L1 cache	64KB, 2-way
per wavelength	4Gb/s	per core	64-byte line
Number of wavelengths	64	L2 cache per core	256KB, 16-way 128-byte line
Protocol	Snoopy MESI protocol	L2 access latency	6 cycles
Protocol msg size	32 bits	Processor	Alpha 21264

Our experimental setup models a 64-core chip-multiprocessor (CMP) on M5 full system simulator. We integrate Wattch EV6 power model to generate the power traces of the system, later we use ISAC [37] for thermal analysis to generate the thermal traces. In our simulator we model the on-chip network as an optical antenna-based broadcast and mesh silicon photonic network similar to what has been proposed in [2] for snoopy cache coherence protocol, where the antenna-based network provides a broadcast network for latency-critical multicast communication and arbitration for resources, while the mesh network provides point to point high-throughput communication. The system and network parameters are specified in Table V as defined in [2], which is a projection of future many-core designs parameters at 64 core assuming 3-D heterogeneous stacking [38] of memory, network, and processing elements would be feasible. We simulated the system under nine different workloads from PARSEC [39], SPLASH2 [29], and SPEComp [40]. The spatial variations in the system varies from 2° to 15° and the temporal variations ranges from 37° to 56°. As for process variations, we assume a 1.2 nm variations across the chip according to our measurements.

A. Power Evaluation

The proposed solution is much more efficient compared to existing solutions to provide reliable communication. In this experiment, we use power overhead as the main criteria for comparison as it provides insight into the overhead of the different approaches. The overhead in system-level techniques presented in this section are mainly two sources. 1) Heater power: microheaters are used to tune the microrings inside a local area. We assume a 0.2 mW/nm power overhead to heat the area of a single microring area to counteract variations by using the low power heating approach in [18]. 2) Voltage tuning power: we have modeled a microring that can have up to 1.65 nm passband shift (this range has been demonstrated experimentally by Manipatruni *et al.* [16]) which is enough for modulation and counteracting variations. We assume a variation of 1.48 nm (equivalent to one channel passband width or 13° variations). The electrical power overhead for variation-tuning is 254 μW in the worst case. We assume for simplicity a linear power overhead per nanometer shift due

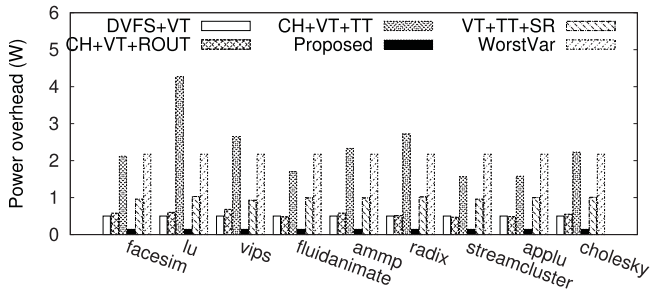

Fig. 9. Power evaluation for different workload.

to the small power overhead. Meanwhile, the optical power overhead is 120 mW for tuning 1.48 nm across the whole on-chip network. We have conducted a simulation-based experiment to compare our solution to the following four system-level techniques in terms of power.

*1) DVFS, Workload Migration, and Voltage Tuning (*DVFS+VT *in Fig. 9) [7]:* In this approach workload migration and DVFS are leveraged to achieve a thermal profile that counteracts process variations leading to a net system-level variations of zero nm. However, workload migration and DVFS cannot achieve exact thermal profile, the variations in thermal profile can be as large as 5° [33]. Moreover, workload migration and DVFS works at the core-level and cannot counteract intracore process variations. Hence, voltage tuning is necessary to align the passband of senders and receivers and compensate any additional variations. To model this approach, we assume a variation of 5° for DVFS inaccuracy. Workload migration has minimal power overhead; hence it is ignored. In addition, we do not account for DVFS power as part of the power overhead since modern systems include DVFS for thermal management and future systems are expected to include different voltage and frequency islands [41]. This approach merely leverages DVFS and does not add any additional circuitry to the system; consequently, it is not an overhead. However, the voltage tuning power (electrical and optical) is an overhead to the system and we account for it in our analysis.

*2) Channel Hopping, Voltage Tuning, and Routing (*CH+VT+ROUT *in Fig. 9) [6]):* In this approach channel hopping is leveraged for coarse-grain matching and voltage tuning for fine-grain tuning. In addition, run-time routing is used to maximize throughput and minimize crosstalk. The power overhead to the system is contributed by voltage tuning, where the tuning range is less than 1.48 nm for a 64 channel system. In our analysis we account for the voltage tuning power (both electrical and optical) overhead as the main power overhead. In addition there is the additional run-time logic and tuning overhead.

*3) Channel Hopping, Voltage Tuning, and Thermal Tuning (*CH+VT+TT *in Fig. 9) [10]:* This approach is similar to *CH+VT+ROUT* (channel hopping, voltage tuning, and routing), the main difference is that voltage tuning is leveraged for blue shift and thermal tuning is used for red-shift in resonant wavelength; thus, reducing the tuning range of voltage tuning and reducing its optical loss at the expense of additional electrical power contributed by thermal heaters. Moreover, it does not leverage any run-time routing for crosstalk nor congestion reduction. The thermal heaters heat up large areas instead of individual microrings which makes the power dependent on area.

*4) Channel Hopping, Thermal Tuning, and Spare Micro-Rings (*CH+TT+SR *in Fig. 9) [11]:* This approach is similar to *CH+VT+TT* (channel hopping, voltage tuning, and thermal tuning); however, the main difference is thermal tuning per micro-ring is used which reduces the area that is being heated and the total heating power. Moreover, this technique adds additional micro-rings at different parts of the spectrum to improve the chances of channel overlap and counteract process variations.

*5) Worst Case Variations (*WorstVar *in Fig. 9):* This approach is similar to our solution using athermal polymer to eliminate thermal variations, and voltage tuning along with channel hopping to overcome process variations. However, in this approach we consider worst case variations instead of two standard deviations to show the impact of statistical variation analysis on our power. We assume a 7.1 nm worst case waveguide width variations [13] and 20 nm worst case silicon thickness variations as guaranteed by SOITEC [12]. According to our simulation 1 nm width variations of a waveguide leads to 0.93 nm shift in resonant wavelength and a 1 nm shift in silicon thickness leads to 1.53 nm shift in resonant wavelength. For worst case width and thickness variations, the resonant shift can be as large as 37.2 nm; thus increasing the total power by 15 times over our proposed solution.

Fig. 9 depicts the power overhead of the four techniques under study compared to the proposed solution. The proposed solution is the most power-efficient one with 0.145 W power overhead, while the other four have power levels within 2 W on average (and 5 W worst case). *DVFS+VT* (DVFS and voltage tuning) having lowest power level with an average of 0.501 W. This is mainly because DVFS is leveraged without introducing additional power overhead, while voltage tuning has shorter tuning range; thus, lowest power level. *CH+VT+ROUT* (channel hopping, voltage tuning, and routing) has an average power overhead of around 0.549 W the second lowest since it leverages voltage tuning for short range tuning (less than 1.48 nm). *CH+TT+SR* has the third highest level power due to the extra microrings used. On average it consumes 0.988 W. WorstVar (worst case variations) has the fourth largest power levels reaching 2.25 W due to worst case design approach that over-provisions power and area. Finally, *CH+VT+TT* (channel hopping, voltage tuning, and thermal tuning) has the highest power level with an average power overhead around 2.35 W because it heats a large area which incurs a high power overhead. All techniques (except for *CH+VT+TT* and *WorstVar*) consume less than 14% of the total power for the silicon photonic network which is small. The *CH+VT+TT* and *WorstVar* consume around 30% of the total power. Compared to our solution, these approaches have a significantly higher power overhead reaching 5× greater on average than the power overhead of the proposed solution. The proposed solution consumes 0.145 W which less than 2% of the total network power. In addition it exhibits very low power overhead thanks to the small tuning range for the

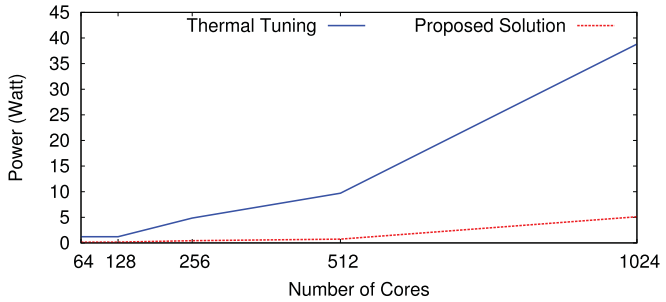

Fig. 10. Power scalability for future many-core systems.

athermal polymer that eliminates any run-time overhead and the voltage tuning for short range.

B. Scalability Analysis

In this section, we show the scalability of different approaches, and find the proposed solution feasible for future many-core system. We make the following assumptions about the process and thermal variations in the future. In many-core system we envision that process variations will improve because the dimensions of silicon photonic devices do not scale with technology; meanwhile, the minimum feature dimension is decreasing; hence, process variations will be less serious. On the other hand, spatial thermal variations are expected to maintain its high level because spatial variations are correlated with maximum operating temperature [42] and the maximum operating temperature will not decrease nor increase according to ITRS projections due to the performance demand and thermal reliability constraints of future systems [41]. From communication demand perspective, to supply enough bandwidth to the network, the number of channels in the network will increase and the channel passband bandwidth will accordingly decrease. Next, we evaluate the different system-level reliability techniques in light of this analysis.

1) Performance: The performance overhead in the different reliability management techniques is small except for the routing algorithm in *CH+VT+ROUT* (channel hopping, voltage tuning, and routing) [6] which would have considerable run-time overhead in the kernel at this large scale system. Assuming the running time of scales linearly with number of cores as Mohamed *et al.*'s study [6] indicates, then the running time overhead for a thousand cores will be 1.92 ms running during the kernel time which is almost 20% of the processor time (assuming a 10 ms time slice per process). However, the role of the algorithm is to reduce crosstalk and improve latency and it is not essential for the network and system reliability; hence, the network will operate reliably without the routing algorithm.

2) Power: The main source of power in the discussed solutions is the thermal tuning power which can be significant especially with large channel count and dynamic thermal profiles. Hence, herein we discuss our proposed solution compared to thermal tuning combined with channel hopping to assess the scalability of our solution compared to the scalability of the other techniques that rely on thermal tuning.

In general, power scales based mainly on number of cores. The limited area of the chip suggests that space division multiplexing (adding more waveguides) is not the solution, instead, WDM is the answer. In this paper we scale the all-optical network suggested by Li *et al.* [2] and compute the overhead of reliability management to compensate variations. We compare two techniques: Thermal tuning and channel hopping (widely used in optical network design [4]) and our proposed reliability management technique. We use a multiprocess workload of a 64-thread of "lu" running on our cores. Next, we assume that number of channels in the optical network and data rate (data rate does not impact the tuning power since temperature changes at the rate of milliseconds) double every two technology nodes to keep up with the bandwidth demand. Moreover, we extend the wavelength range of channels to 350 nm, which is the 3 dB of the silicon waveguide we designed. We also account in channel bandwidth reduction due to increase of channel count. The total tuning power due to thermal tuning reaches 38 W at 1024 cores consuming more than 25% of the total power consumption of the system (which is limited to 150 W due to thermal constraints [41]). On the other hand, our approach reaches 5 W at 1024 cores as shown in Fig. 10 which is low enough for power efficient operation.

Our proposed solution, compared to other reliability management solutions, has no run-time control overhead; consequently, there is no performance overhead and no additional circuitry needed. Moreover, the voltage tuning power overhead is decreasing since the channel bandwidth is decreasing and process variations are improving.

VI. Conclusion

The barrier to wide scale deployment of silicon photonics is reliability. Process variations and run-time temperature fluctuations affect silicon photonic communication system reliability. We present a new reliability-aware design flow to provide reliable on-chip communication in the many-core systems. This flow is based on abstracting the system into two levels: The device- and the network-level. The flow proceeds in two phases, an analysis phase which drives the design phase. The outcome is a detailed design of a reliability management solution. Based on our flow, we propose a novel system-level solution that can be coupled to variation-aware design to produce reliable optical interconnection networks at the on-chip many-core scale. Evaluation results show high reliability, low run-time overhead, low power consumption, and scalability compared to other existing techniques.

References

[1] R. G. Beausoleil, J. Ahn, N. Binkert, A. Davis, D. Fattal, M. Fiorentino, *et al.*, "A nanophotonic interconnect for high-performance many-core computation," in *Proc. 16th IEEE Symp. High Perform. Interconnects*, Aug. 2008, pp. 182–189.

[2] Z. Li, M. Mohamed, X. Chen, H. Zhou, A. Mickelson, L. Shang, *et al.*, "Iris: A hybrid nanophotonic network design for high-performance and low-power on-chip communication," *J. Emerg. Technol. Comput. Syst.*, vol. 7, pp. 8:1–8:22, Jul. 2011.

[3] M. Petracca, B. G. Lee, K. Bergman, and L. P. Carloni, "Design exploration of optical interconnection networks for chip multiprocessors," in *Proc. 16th IEEE Symp. High Perform. Interconnects*, Aug. 2008, pp. 31–40.

[4] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, *et al.*, "Corona: System implications of emerging nanophotonic technology," in *Proc. 35th ISCA*, 2008, pp. 153–164.

[5] X. Chen, M. Mohamed, Z. Li, L. Shang, and A. Mickelson, "Process and thermal variation in silicon photonic devices," *Appl. Opt.*, 2013, to be published.

[6] M. Mohamed, Z. Li, X. Chen, L. Shang, A. R. Mickelson, M. Vachharajani, *et al.*, "Power-efficient variation-aware photonic on-chip network management," in *Proc. ACM/IEEE ISLPED*, Aug. 2010, pp. 31–36.

[7] Z. Li, M. Mohamed, X. Chen, E. Dudley, K. Meng, L. Shang, *et al.*, "Reliability modeling and management of nanophotonic on-chip networks," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 20, no. 1, pp. 98–111, Jan. 2012.

[8] Z. Li, M. Mohamed, X. Chen, A. Mickelson, and L. Shang, "Device modeling and system simulation of nanophotonic on-chip networks for reliability, power and performance," in *Proc. 48th Design Autom. Conf.*, 2011, pp. 735–740.

[9] V. Raghunathan, W. N. Ye, J. Hu, T. Izuhara, J. Michel, and L. Kimerling, "Athermal operation of silicon waveguides: Spectral, second order and footprint dependencies," *Opt. Exp.*, vol. 18, no. 17, pp. 17631–17639, Aug. 2010.

[10] C. Nitta, M. Farrens, and V. Akella, "Addressing system-level trimming issues in on-chip nanophotonic networks," in *Proc. IEEE 17th Int. Symp. HPCA*, Feb. 2011, pp. 122–131.

[11] Y. Xu, J. Yang, and R. Melhem, "Tolerating process variations in nanophotonic on-chip networks," in *Proc. 39th Annu. ISCA*, Jun. 2012, pp. 142–152.

[12] (2008). *ePIXfab Website* [Online]. Available: http://www.epixfab.eu/

[13] S. K. Selvaraja, W. Bogaerts, P. Dumon, D. Van Thourhout, and R. Baets, "Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using cmos fabrication technology," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, no. 1, pp. 316–324, Jan./Feb. 2010.

[14] X. Chen, Z. Li, M. Mohamed, L. Shang, and A. Mickelson, "Matrix analysis of nanophotonic devices," in *Proc. Int. Conf. Fiber Opt. Photon.*, Dec. 2010.

[15] M. Mohamed, Z. Li, X. Chen, A. Mickelson, and L. Shang, "Modeling and analysis of micro-ring based silicon photonic interconnect for embedded systems," in *Proc. 7th IEEE/ACM/IFIP Int. Conf. Hardw./Softw. Codesign Syst. Synthesis*, Oct. 2011, pp. 227–236.

[16] S. Manipatruni, R. K. Dokania, B. Schmidt, N. Sherwood-Droz, C. B. Poitras, A. B. Apsel, *et al.*, "Wide temperature range operation of micrometer-scale silicon electro-optic modulators," *Opt. Lett.*, vol. 33, no. 19, pp. 2185–2187, 2008.

[17] L. Chen, N. Sherwood-Droz, and M. Lipson, "Compact bandwidth tunable microring resonators," in *Proc. Conf. Lasers Electro-Opt./Quantum Electron. Laser Sci. Conf. Photon. Appl. Syst. Technol.*, 2008, pp. 1–3.

[18] P. Dong, W. Qian, H. Liang, R. Shafiiha, D. Feng, G. Li, *et al.*, "Thermally tunable silicon racetrack resonators with ultralow tuning power," *Opt. Exp.*, vol. 18, no. 19, pp. 20298–20304, Sep. 2010.

[19] L. Zhou, K. Kashiwagi, K. Okamoto, R. Scott, N. Fontaine, D. Ding, *et al.*, "Towards athermal optically-interconnected computing system using slotted silicon microring resonators and RF-photonic comb generation," *Appl. Phys. A, Mater. Sci. Process.*, vol. 95, no. 4, pp. 1101–1109, Jun. 2009.

[20] X. Han, M. Zhao, J. Zhang, L. Wang, J. Teng, J. Wang, *et al.*, "Design of athermal all-polymer waveguide microring resonator," in *Proc. Asia Commun. Photon. Conf. Exhibit.*, 2009, no. ThN5, pp. 1–3.

[21] B. Guha, B. B. C. Kyotoku, and M. Lipson, "CMOS-compatible athermal silicon microring resonators," *Opt. Exp.*, vol. 18, no. 4, pp. 3487–3493, 2010.

[22] N. Kobayashi, N. Zaizen, and Y. Kokubun, "Athermal and polarization-independent microring resonator filter using stress control," *Jpn. J. Appl. Phys.*, vol. 46, pp. 5465–5469, Aug. 2007.

[23] U. Levy, K. Campbell, A. Groisman, S. Mookherjea, and Y. Fainman, "On-chip microfluidic tuning of an microring resonator," in *Proc. Conf. Lasers Electro-Opt./Quantum Electron. Laser Sci. Conf. Photon. Appl. Syst. Technol.*, 2006, no. CFC1, pp. 1–2.

[24] P. T. Rakich, M. A. Popovic, M. R. Watts, T. Barwicz, H. I. Smith, and E. P. Ippen, "Ultrawide tuning of photonic microcavities via evanescent field perturbation," *Opt. Lett.*, vol. 31, no. 9, pp. 1241–1243, May 2006.

[25] D. Espinoza, X. Chen, M. Mohamed, H. Zhou, E. Dudley, W. Park, *et al.*, "Nanometric polymer coatings for silicon on insulator circuits," *Proc. SPIE*, vol. 8173, pp. 81730H-1–81730H-10, Aug. 2011.

[26] L. Zhou, K. Okamoto, and S. Yoo, "Athermalizing and trimming of slotted silicon microring resonators with uv-sensitive PMMA upper-cladding," *IEEE Photon. Technol. Lett.*, vol. 21, no. 17, pp. 1175–1177, Sep. 1, 2009.

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

[28] Y. Xu, J. Yang, and R. Melhem, "Channel borrowing: An energy-efficient nanophotonic crossbar architecture with light-weight arbitration," in *Proc. 26th ACM Int. Conf. Supercomput.*, 2012, pp. 133–142.

[29] *SPLASH2 Website* [Online]. Available: http://www-flash.stanford.edu/apps/SPLASH/

[30] G. Kurian, J. E. Miller, J. Psota, J. Eastep, J. Liu, J. Michel, *et al.*, "ATAC: A 1000-core cache-coherent processor with on-chip optical network," in *Proc. 19th Int. Conf. Parallel Archit. Compil. Tech.*, 2010, pp. 477–488.

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

[32] C. Nitta, M. Farrens, and V. Akella, "Resilient microring resonator based photonic networks," in *Proc. 44th Annu. IEEE/ACM Int. Symp. Microarchit.*, Dec. 2011, pp. 95–104.

[33] H. Hanson, S. W. Keckler, S. Ghiasi, K. Rajamani, F. Rawson, and J. Rubio, "Thermal response to DVFS: Analysis with an intel pentium M," in *Proc. Int. Symp. Low Power Electron. Design*, 2007, pp. 219–224.

[34] D. Oh, C. Chen, N. S. Kim, and Y. H. Hu, "The compatibility analysis of thread migration and DVFS in multi-core processor," in *Proc. 11th ISQED*, Mar. 2010, pp. 866–871.

[35] N. L. Binkert, R. G. Dreslinski, L. R. Hsu, K. T. Lim, A. G. Saidi, and S. K. Reinhardt, "The M5 simulator: Modeling networked systems," *IEEE Micro*, vol. 26, no. 4, pp. 52–60, Jul./Aug. 2006.

[36] D. Brooks, V. Tiwari, and M. Martonosi, "Wattch: A framework for architectural-level power analysis and optimizations," in *Proc. 27th Int. Symp. Comput. Archit.*, 2000, pp. 83–94.

[37] Y. Yang, C. Zhu, L. Shang, and R. P. Dick. *Incremental Self-Adaptive Chip-Package Thermal Analysis Software* http://post.queensu.ca/~shangl/isac/index.html and http://robertdick.org/tools.html

[38] B. Black, M. Annavaram, N. Brekelbaum, J. DeVale, L. Jiang, G. H. Loh, *et al.*, "Die stacking (3D) microarchitecture," in *Proc. 39th Annu. IEEE/ACM Int. Symp. Microarchit.*, Dec. 2006, pp. 469–479.

[39] C. Bienia, S. Kumar, J. P. Singh, and K. Li, "The PARSEC benchmark suite: Characterization and architectural implications," Oct. 2008 [Online]. Available: https://www.cs.princeton.edu/research/techreps/TR-811-08

[40] V. Aslot, M. Domeika, R. Eigenmann, G. Gaertner, W. B. Jones, and B. Parady, "SPEComp: A new benchmark suite for measuring parallel computer performance," in *Proc. Workshop OpenMP Appl. Tools*, 2001, pp. 1–10.

[41] (2009). *International Technology Roadmap for Semiconductors (ITRS)* [Online]. Available: http://www.itrs.net/Links/2009ITRS/Home2009.htm

[42] M. Cho, N. Sathe, M. Gupta, S. Kumar, S. Yalamanchilli, and S. Mukhopadhyay, "Proactive power migration to reduce maximum value and spatiotemporal non-uniformity of on-chip temperature distribution in homogeneous many-core processors," in *Proc. 26th Annu. IEEE SEMI-THERM*, Feb. 2010, pp. 180–186.

Moustafa Mohamed (S'10) received the Ph.D. degree from the University of Colorado at Boulder, Boulder, CO, USA, in 2013.

Zheng Li (S'07) was a Research Associate with the University of Colorado at Boulder, Boulder, CO, USA.

Xi Chen (S'07) is currently pursuing the Ph.D. degree in electrical, computer, and energy engineering with the University of Colorado at Boulder, Boulder, CO, USA.

Li Shang (S'99–M'04) is an Associate Professor with the Department of Electrical, Computer, and Energy Engineering, University Colorado at Boulder, Boulder, CO, USA.

Alan R. Mickelson (S'72–M'78–SM'92) is an Associate Professor with the Department of Electrical, Computer, and Energy Engineering, University of Colorado at Boulder, Boulder, CO, USA.